

Peer Reviewed Paper **openaccess** [Special Issue on Chemometrics in Hyperspectral Imaging](#)

Near infrared hyperspectral images and pattern recognition techniques used to identify etiological agents of cotton anthracnose and ramulosis

Priscila S.R. Aires,^a Francisco F. Gambarra-Neto,^b Wirton M. Coutinho,^c Alderi E. Araujo,^d Gilvan Ferreira da Silva,^e Josivanda P.G. Gouveia^f and Everaldo P. Medeiros^{g,*}

^aFederal University of Campina Grande, Post-graduate Program in Agricultural Engineering, CEP 58.429-140, Campina Grande-PB, Brazil. ORCID: <https://orcid.org/0000-0003-0991-2068>

^bFederal University of Paraíba, Graduate Program in Agronomy, CEP 58.397-000, Areia, PB, Brazil. ORCID: <https://orcid.org/0000-0002-8800-5644>

^cEmbrapa Algodão, CEP 58428-095, Campina Grande, PB, Brazil. ORCID: <https://orcid.org/0000-0001-5159-7829>

^dEmbrapa Algodão, CEP 58428-095, Campina Grande, PB, Brazil. ORCID: <https://orcid.org/0000-0003-0257-1539>

^eEmbrapa Amazônia Ocidental, CEP 69010-970, Manaus, AM, Brazil. ORCID: <https://orcid.org/0000-0003-2828-8299>

^fFederal University of Campina Grande, Post-graduate Program in Agricultural Engineering, CEP 58.429-140, Campina Grande-PB, Brazil. ORCID: <https://orcid.org/0000-0002-2047-986X>

^gEmbrapa Algodão, CEP 58428-095, Campina Grande, PB, Brazil. ORCID: <https://orcid.org/0000-0001-7204-3040>.
E-mail: everaldo.medeiros@embrapa.br

Hyperspectral imaging near infrared (HSI-NIR) has the potential to be used as a non-destructive approach for the analysis of new microbiological matrices of agriculture interest. This article describes a new method for accurately and rapidly classifying the etiological agents *Colletotrichum gossypii* (CG) and *C. gossypii* var. *cephalosporioides* (CGC) grown in a culture medium, using scattering reflectance HSI-NIR and multivariate pattern recognition analysis. Five strains of CG and 46 strains of CGC were used. CG and CGC strains were grown on Czapek-agar medium at 25 °C under a 12-hour photoperiod for 15 days. Molecular identification was performed as a reference for the CG and CGC classes by polymerase chain reaction of the intergenic spacer region of rDNA. The scattering coefficient μ_s and the absorption coefficient μ_a were obtained, which resulted in a μ_s value for CG of 1.37×10^{19} and for CGC of 5.83×10^{-11} . These results showed that the use of the standard normal variate was no longer essential and reduced the spectral range from 1000–2500 nm to 1000–1381 nm. The results evidenced two type II errors for the CG 457-2 and CGC 39 samples in the soft independent modelling model of the analogy model. There were no classification errors using the algorithm of the successive projections for variable selection in linear discriminant analysis (SPA-LDA). A parallel validation of the results obtained with SPA-LDA was performed using a box plot analysis with the 11 variables selected by SPA, in which there were no outliers for the HSI-NIR models. The new HSI-NIR and SPA-LDA procedures for the classification of CG and CGC etiological agents are noted for their greater analytical speed, accuracy, simplicity, lower cost and non-destructive nature.

Keywords: fungal identification, fungal taxonomy, non-destructive analysis, cotton crop, hyperspectral image

Correspondence

Everaldo P. Medeiros (everaldo.medeiros@embrapa.br)

Received: 29 September 2017

Revised: 10 February 2018

Accepted: 26 February 2018

Publication: 24 April 2018

doi: 10.1255/jsi.2018.a8

ISSN: 2040-4565

Citation

P.S.R. Aires et al., "Near infrared hyperspectral images and pattern recognition techniques used to identify etiological agents of cotton anthracnose and ramulosis", *J. Spectral Imaging* 7, a8 (2018). <https://doi.org/10.1255/jsi.2018.a8>

© 2018 The Authors

This licence permits you and anybody else to use, share, copy and redistribute the paper in any medium or any format provided that a full citation to the original paper in this journal is given and the use is not for commercial purposes.



Introduction

Cotton ramulosis and anthracnose are distinct diseases caused by pathogens that have identical morphological characteristics.¹ The ramulosis pathogen causes death of the apical meristem of the cotton plant, which results in stunting, the shortening of internodes and the excessive development of branches with witches' broom symptoms. This disease was first described in Brazil in 1937, and its etiological agent was identified as *Colletotrichum gossypii* var. *cephalosporioides* (CGC).^{2,3} This variant was considered more aggressive than *Colletotrichum gossypii* (CG), previously described in the United States of America, causing seedling death as well as cotton leaf and boll lesions.^{4,5}

The CG and CGC pathogens are transmitted to adult cotton plant through seed.^{6,7} Only CGC, however, has quarantine importance in Brazil. The Brazilian phytosanitary authority has proposed a zero tolerance level for ramulosis pathogen on commercial cotton seeds.⁸

The identification of CG and CGC using traditional methods is difficult, even for experienced professionals. Tanaka *et al.*⁶ proposed the separation of these fungi based on morphological markers of fungal structures growing on cotton seeds submitted to seed blotter tests, while Silva-Mann *et al.*⁹ suggested differentiation by amplified fragment length polymorphism molecular markers. However, according to Hyde *et al.*,¹⁰ there are no consistent reports concerning morphological and molecular markers used to differentiate the two varieties. Some authors have considered *C. gossypii* to be a synonym of *C. gloeosporioides*, but this taxon was not accepted by Sutton.¹¹ Recently, a molecular phylogenetic analysis conducted by Salustiano *et al.*¹² showed that *C. gossypii* and *C. gossypii* var. *cephalosporioides* formed distinct lineages inside the clade of *Colletotrichum theobromicola*, a member of the *C. gloeosporioides* species complex, which allows distinction between the ramulosis and anthracnose pathogens of cotton.

Molecular techniques have been successfully used in the differentiation of species in a wide range of genera of pathogenic fungi, including *Colletotrichum*.^{9,12,13} These techniques, however, are complex, slow, laborious and costly compared to spectral techniques, such as hyperspectral imaging near infrared (HSI-NIR) technology. HSI-NIR is an analytical tool with broad analytical capability for the verification and quantification of analytical compounds in biological systems.¹⁴ HSI-NIR is an excellent rapid and low-cost alternative fungi identification

technique when compared to DNA-based methods.¹⁵ The HSI-NIR technique considers the spectral and spatial distributions of the sample at the pixel level, in that each pixel will correspond to one continuum spectrum with more than a hundred wavelengths or channels.^{15,16} HSI-NIR is a physical, non-destructive technique that requires minimal or no sample preparation.^{17,18}

HSI-NIR technology, which includes both the instrumental and data treatment features involving chemometrics, has been successfully used to detect mycotoxins in agricultural products, such as wheat kernels,¹⁹ maize²⁰ and barley.²¹ The technique has also been used to detect and identify *Fusarium verticillioides* in maize²² and the storage of aflatoxicogenic fungi associated with wheat.²³ Hyperspectral imaging techniques were also successfully used to discriminate citrus canker from other factors in confounding diseases.²⁴

The approach in the treatment of HSI data involves the use of chemometric techniques combined or developed to meet specific needs.²⁵ Among the various multivariate analysis strategies, pattern recognition techniques have implementation advantages in highly complex matrices employing HSI-NIR.^{26,27} Classification applications using HSI and PLS-DA have received considerable attention because they are easy to implement and provide better results.²⁸

However, few studies have been reported using variable selection criteria in HSI data for pattern recognition. In a recent study by Calvini *et al.*,²⁹ the transference of classification models with the selection of *a priori* variables in spectral bands was carried out. In this context, strategies based on dedicated algorithms for variable selection and discriminant analysis (SPA-LDA) may have the advantage of improving computational capacity and simplifying data processing.³⁰

Theory

HSI-NIR techniques are used on the identification of chemical bonds based on their molecular vibrations, generating a spectrum for each pixel that forms the digital image.³¹ Therefore, the hypercube of an image generates information of high complexity,³² which, when associated with multivariate analysis methods for pattern recognition and classification, may be useful for the identification and separation of fungal samples.

White and dark colours were used as references for the calibration of the hyperspectrometer. The dark current effect is removed by the dark reference of charge-coupled device (CCD)³³ detectors, which are sensitive to thermal changes. The calibration of the image (R) was defined by Equation 1:

$$R = \frac{R_0 - D}{W - D} \times 100 \quad (1)$$

where R_0 is the hyperspectral image, D is the dark image with zero reflectance and W is the white reference image taken from a standard white Teflon[®] reference plate with 99% reflectance.^{33,34}

Light propagation in a turbid medium is modelled with the radiative transfer equation, which neglects the wave properties and considers the energy transport. The light propagation can be numerically approximated as a scattering process when the spread is dominant over the absorption as a continuous beam of infinitely small diameter vertically striking a scattering turbid medium. This process describes the diffuse reflectance as scattering coefficient³⁵ and absorption coefficient.³⁵ The radial distance r in cm measured from the centre of the illumination beam (Equation 2):

$$R(r) = \frac{a'}{4\pi} \left[\frac{1}{\mu'_t} \left(\mu_{eff} + \frac{1}{r_1} \right) \frac{\exp(-\mu_{eff} r_1)}{r_1^2} + \left(\frac{1}{\mu'_t} + \frac{4A}{3\mu'_t} \right) \left(\mu_{eff} + \frac{1}{r_2} \right) \frac{\exp(-\mu_{eff} r_2)}{r_2^2} \right] \quad (2)$$

where a' is the transport reflection coefficient [$a' = \mu_s / (\mu_a + \mu_s)$]; μ'_t is the effective attenuation coefficient ($\mu_{eff} = [3\mu_a + \mu_s]^{1/2}$); r_1 [$r_1 = [(1/\mu'_t)^2 + r^2]^{1/2}$] and r_2 [$r_2 = [(1/\mu'_t)^2 + 4^a/3\mu'_t)^2 + r^2]^{1/2}$] are the radial distances; and A is an internal reflection coefficient that depends on the medium.

The unpaired boundary condition or the internal reflection coefficient can be estimated with the following empirical equation:

$$A = \frac{1 + r_d}{1 - r_d} \quad (3)$$

in which r_d is determined according to the following approximation:

$$r_d \approx 1.44n_r^{-2} + 0.710n_r^{-1} + 0.668 + 0.0636n_r \quad (4)$$

in which n_r is the relative reflectance index of the sample.³⁴

When light scattering dominates over absorption ($\mu_s \gg \mu_a$), the model of Farrell *et al.*³⁴ or the model of

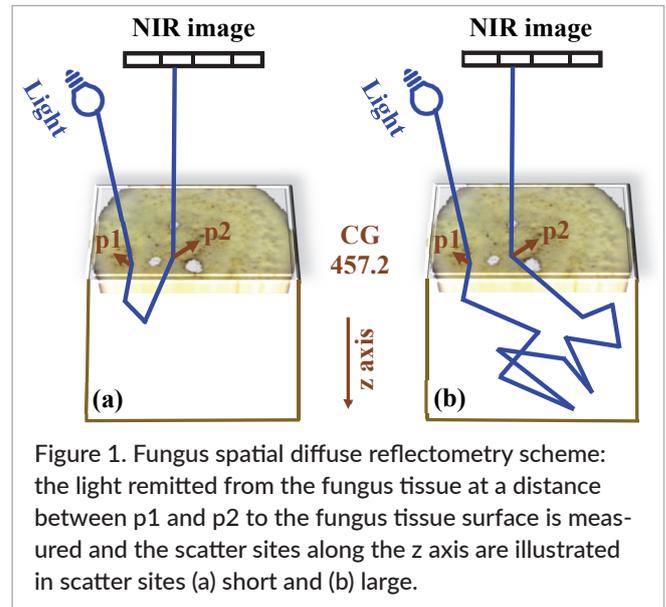


Figure 1. Fungus spatial diffuse reflectometry scheme: the light remitted from the fungus tissue at a distance between p1 and p2 to the fungus tissue surface is measured and the scatter sites along the z axis are illustrated in scatter sites (a) short and (b) large.

Michels *et al.*³⁵ of the propagation of light based on scattering approximation can be used to describe the relationship between optical properties and spatially resolved reflectance profiles acquired with the HSI-NIR scattering configuration.

The steady-state scattering theory model describes the diffuse dependence of radial tissues light reflectance (Figure 1), with an accuracy of 0.5 mm for radial distances.³⁴

Diffuse optical reflection is used to reconstruct absorption images from diffuse light continuous wave measurements. The image algorithm allows them to be reconstructed from measurements made with a single light source.³⁶ The diffused-reflected light images are recorded by a CCD camera in turbid medium containing an absorbent lesion,^{37,38} including NIR spectroscopy images.³⁹

Diffuse optical imaging is the use of NIR light measurements with scattering techniques applied to modelling and reconstructing images.⁴⁰ This imaging examines absorption as well as the scattering properties of biological tissues, translation and access to a variety of inaccessible physiological parameters.⁴¹ An example is fungi which show non-homogeneity.⁴² Hyperspectral fungi images are composed by waves proliferated by random means. The transport of multiple scattered waves is accounted for by the scattering theory that derives from a treatment of the diffuse intensity that includes the effects of the boundary layers, where mesoscopic correlations provide multiple scattering intensities. These correlations are the result of underlying wave characters.⁴³

Scattering theory for infinite, homogeneous and turbid media describes the nature of the propagation.⁴⁴ Dynamic light scattering is routinely used to measure the scattering of the surfactant mycelium and provides mutual scattering coefficients for all of the components of the surfactant.⁴⁵ The scattering coefficient for light transport in scattering and absorbing media presents a novel derivation of the scattering coefficient, which is restricted neither to low absorption nor to a situation in which the specific intensity is quasi-isotropic.⁴⁶ Structural height of the diffuser is correlated to the scattering coefficient's frequency characteristics, while the diffuser coverage density is correlated to the values of the scattering coefficients.⁴⁷ At the limit of an optically thick medium in which the mean free path of each photon is much smaller than the average size, the multiple scattering effects can be approximated by a scattering process.^{48,49}

The absorbance signals ratio contains information on the scattering coefficients of sodium azide, benzoic acid, amino acids, peptides and proteins.⁵⁰ The CO₂ and O₂ gases scattering coefficients are dependent on the concentration of the *Aspergillus niger* fermenter fungus biomass.⁵¹ The applicability of scattering theory for the determination of tissue optical properties from steady-state diffuse reflectance was investigated through the analytical expressions from scattering theory.⁵² The scattering theory model was developed with an HSI-NIR instrument to obtain CG and CGC fungi hypercubes.

The fungi presented fat emulsions whose scattering coefficient (μ_s) and absorption coefficient (μ_a) calculations were performed according to Michels *et al.*³⁵ to identify the existence of the scattering and to identify the NIR hyperspectral region that presented the greater difference between CG and CGC fungi.

The objective of this study was to identify and separate strains of cotton anthracnose fungus (CG) and cotton ramulosis fungus (CGC) using the HSI-NIR technique. We investigated light propagation models to verify the configuration of the hyperspectral scattering image and classification models with soft independent modelling of class analogy (SIMCA) and the selection of variables by the successive projections for linear discriminant analysis (SPA-LDA) algorithm. A number of studies have demonstrated the feasibility of this technique in microbiological applications. For the identification of CG and CGC strains, which are morphologically identical, we tested the HSI-NIR technique using tools for the selection of spectral variables in the classification of the two species.

This study is the first to our knowledge involving this technique.

Material and methods

Cultivation of fungal strains

In this study, five CG strains and 46 CGC strains were used. All the strains belong to the Embrapa Algodão Collection of Phytopathogenic Microorganisms. CG and CGC strains were originally isolated from diseased cotton plants or seeds with typical symptoms of anthracnose or ramulosis, collected from different cotton-producing areas in the Brazilian states of Goiás and Mato Grosso (Table 1).

CG and CGC strains were grown in 90-mm Petri dishes containing 20 mL of Czapek-agar (3 % w/v sucrose, 0.3 % w/v NaNO₃, 0.1 % w/v K₂HPO₄, 0.05 % w/v MgSO₄•7H₂O, 0.05 % w/v KCl and 0.001 % FeSO₄•7H₂O w/v) solidified with 1.5 w/v agar and incubated for 15 days in a growth chamber at a temperature of 25 °C for a photoperiod of 12 hours.

Pathogenicity tests

The pathogenicity tests were conducted by inoculating three CG strains (CNPA 0735, CNPA 0736 and CNPA 0738) and five CGC strains (CNPA 0060, CNPA 0080, CNPA 0094, CNPA 0117 and CNPA 0123) on cotton plants (cultivar BRS Cedro, which is susceptible to both pathogens). Each strain was grown in Petri dishes containing PDA medium (20 % v/v potato infusion, 2 % w/v dextrose) solidified with 2 % w/v agar and incubated at 25 °C for a 12-hour photoperiod for 10 days to induce pathogen sporulation. Conidial suspensions were prepared by adding 10 mL of sterile distilled water and three drops of Tween-20 (5 %) to each plate. The suspensions were obtained by carefully brushing the colonies with a sterile loop. The resulting suspensions were then filtered through a sterilised double layer of cheesecloth to obtain pure conidial suspensions and adjusted to 1 × 10⁵ spores per mL using a cell-counting hemacytometer.

Plants of the cotton cultivar BRS Cedro were grown in 5-L pots containing a 3:1 mixture of peat and vermiculite (v/v). Plants were grown in a greenhouse at 25 ± 5 °C and 80% relative humidity. The plants were inoculated 36 days after emergence (DAE) by spraying the plants with the previously prepared conidial suspension with an air-pressurised sprayer until runoff occurred. The control

Table 1. *C. gossypii* (CG) and *C. gossypii* var. *cephalosporioides* (CGC) strains used in this study.

Species	Identification CNPA	Host-substrate	Place of origin
CG	0037	Cotton-seed	Acreúna, GO
CG	0735	Cotton-seed	MT
CG	0736	Cotton-seed	MT
CG	0737	Cotton-seed	MT
CG	0738	Cotton-seed	MT
CGC	0038	Cotton-leaf	Rondonópolis, MT
CGC	0039	Cotton-leaf	Mineiros, GO
CGC	0040	Cotton-leaf	Pedra Preta, MT
CGC	0041	Cotton-leaf	Pedra Preta, MT
CGC	0043	Cotton-leaf	Novo São Joaquim, MT
CGC	0044	Cotton-leaf	Pedra Preta, MT
CGC	0045	Cotton-leaf	Pedra Preta, MT
CGC	0046	Cotton-leaf	Pedra Preta, MT
CGC	0048	Cotton-leaf	Rondonópolis, MT
CGC	0049	Cotton-leaf	Rondonópolis, MT
CGC	0051	Cotton-leaf	Sapezal, MT
CGC	0052	Cotton-leaf	Primavera do Leste, MT
CGC	0053	Cotton-leaf	Santa Helena de Goiás, GO
CGC	0056	Cotton-leaf	Alto Taquari, MT
CGC	0057	Cotton-leaf	Mineiros, GO
CGC	0059	Cotton-leaf	Mineiros, GO
CGC	0061	Cotton-leaf	Primavera do Leste, MT
CGC	0062	Cotton-leaf	Mineiros, GO
CGC	0063	Cotton-leaf	Mineiros, GO
CGC	0067	Cotton-leaf	Chapadão do Céu, GO
CGC	0069	Cotton-leaf	Pedra Preta, MT
CGC	0070	Cotton-leaf	Itiquira, MT
CGC	0071	Cotton-leaf	Itiquira, MT
CGC	0072	Cotton-leaf	Primavera do leste, MT
CGC	0074	Cotton-leaf	Itiquira, MT
CGC	0075	Cotton-leaf	Itiquira, MT
CGC	0078	Cotton-leaf	Primavera do Leste, MT
CGC	0079	Cotton-leaf	Primavera do Leste, MT
CGC	0082	Cotton-leaf	Chapadão do Céu, GO
CGC	0083	Cotton-leaf	Itiquira, MT
CGC	0084	Cotton-leaf	Pedra Preta, MT
CGC	0086	Cotton-leaf	Novo São Joaquim, MT
CGC	0104	Cotton-leaf	Santa Helena de Goiás, GO
CGC	0105	Cotton-leaf	Santa Helena de Goiás, GO
CGC	0107	Cotton-leaf	Santa Helena de Goiás, GO

Table 1 (continued). *C. gossypii* (CG) and *C. gossypii* var. *cephalosporioides* (CGC) strains used in this study.

Species	Identification CNPA	Host-substrate	Place of origin
CGC	0113	Cotton-leaf	Campo Verde, MT
CGC	0116	Cotton-leaf	Campo Verde, MT
CGC	0117	Cotton-leaf	Campo Verde, MT
CGC	0121	Cotton-leaf	Campo Verde, MT
CGC	0123	Cotton-leaf	Santa Helena de Goiás, GO
CGC	0124	Cotton-leaf	Santa Helena de Goiás, GO
CGC	0127	Cotton-leaf	Cachoeira Dourada, GO
CGC	0132	Cotton-leaf	Alto Taquari, MT
CGC	0133	Cotton-leaf	Alto Taquari, MT
CGC	0134	Cotton-leaf	Alto Taquari, MT
CGC	0135	Cotton-leaf	Itiquira, MT

treatment consisted of plants sprayed only with sterile distilled water. Each treatment consisted of four replicates arranged in blocks at random.

Reference assays using molecular identification by the intergenic spacer region of rDNA

Cultivation of fungal strains and DNA extraction
CG and CGC strains were maintained in a culture medium for rapid growth (0.2% peptone w/v, 1% dextrose w/v, 0.15% casein w/v and 0.2% yeast extract). Mycelial mass was grown in 250-mL vials with 50 mL of medium, described above, and agitated at 150 rpm for approximately 3 days at 25 °C.

The total DNA of the strains (Table 1) was extracted following the cetyl trimethylammonium bromide (CTAB) extraction procedures of macerated mycelia in liquid nitrogen.⁵³ Quantification of the DNA extracted was performed using a Nano Drop 2000 spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), with determination of protein concentration by measuring absorbance at 280 nm. In addition, a 0.8% agarose gel was used for capillary electrophoresis analysis. All samples were diluted to a final concentration of 50 ng μL^{-1} .

Molecular diagnosis of GC and CGC

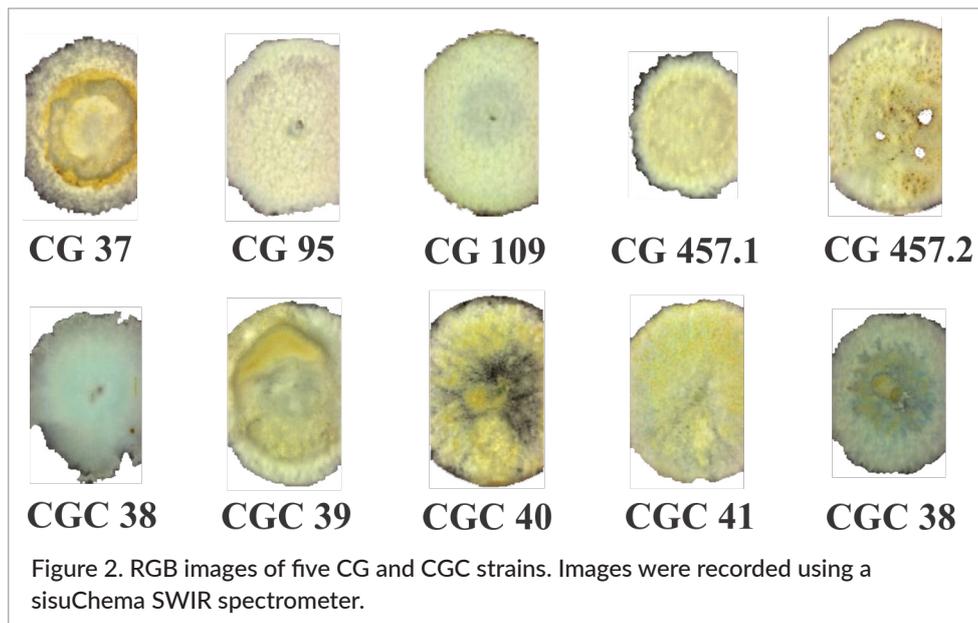
For the differential molecular diagnosis of CG and CGC strains, the intergenic spacer (IGS) region of rDNA was used. A 3.3-kb GC fragment and a 2.7-kb CGC fragment were completely sequenced, and gaps between CG and

CGC were located to design of a set of primers that allowed molecular diagnosis via polymerase chain reaction (PCR).

IGS primers F2-5'GAAAAGTAAGTACCCCCGAA3' and IGS detector R-5'TGGCGGCGGTGAGTCGGGGTGC3' amplified a 432-bp fragment for GC and 148-bp fragment for CGC. PCRs were performed in a volume of 20 μL containing 50 ng of total DNA; buffer solution (pH 8.8) at 25 °C (100 mmol L^{-1} Tris-HCl); 500 mmol L^{-1} aqueous KCl (0.8% v/v); 2.0 mmol L^{-1} MgCl_2 ; 0.4 mmol L^{-1} DNTPs; 1 U of Taq DNA polymerase (DNA Express Cat. No. 0300.0003.0500); and each primer at 0.2 $\mu\text{mol L}^{-1}$. The conditions for amplification included initial denaturation at 94 °C for 3 min followed by 40 denaturation cycles at 94 °C for 15 s, annealing of the primers at 65 °C for 15 s and synthesis at 72 °C for 30 s. After the cycles were completed, an extension was performed at 72 °C for 10 min.

Hyperspectral NIR image

HSI-NIR of the CG and CGC strains was performed using a sisuChema SWIR-Specim[®] spectrometer (Spectral Imaging Ltd, Oulu, Finland) with a 1000–2500 nm spectral range, 6.3 nm spectral sampling, 10 nm resolution, 150 \times 150 μm pixel image size, 256 wavelength bands, a high-performance camera, a 50-mm lens and ChemaDAQ software. The HSI-NIR system was customised with a set of 26 halogen light bulbs (12 V, 10 W and 9.5 mm) as well as a MCT detector (HgCdTe).



The height between the lens and the sample was 31 cm, the velocity of the carrier was set at 30 mm s^{-1} to ensure the same spatial shape of the sample in the image and the exposure time was adjusted to $3000\mu\text{s}$ to ensure an intensity of light suitable for a good image. During the measurements, Petri dishes were opened for the recording of the HSI-NIR images. The HSI-NIR images of five CG and CGC samples analysed are shown in Figure 2.

The acquisition by HSI-NIR includes the digitisation of the linear matrix by the detector along the X axis and the movement of the sample on the Y axis as well as the NIR spectra on the perpendicular Z axis of the sample. Thus, a given 3D hypercube representing the HSI-NIR contains spatial (pixel) and spectral (NIR) image information that can be used to identify the pathogens that are present in CG and CGC samples.

HSI-NIR images were saved in raw format, and data preprocessing was performed from the means of each HSI-NIR sample. The data were centred on the average of the NIR hyperspectra, smoothed in accordance with the Savitzky-Golay⁵⁴ technique with a window of five wavelengths and fit to second-order polynomials. The mean spectra of each strain were obtained by selecting the image of each sample and using *Evince* software version 2.6.0 (*UmBio*, Sweden).

Models of light scattering and light absorption for the HSI-NIR technique

To acquire spectral profiles with the HSI-NIR configuration, spectral properties were estimated for each wave-

length by the functions of the scattering coefficient (μ_s) and absorption coefficient (μ_a) described by Michels *et al.*³⁵ The means and standard deviations of five CG samples and 36 CGC samples were calculated, resulting in only two spectral profiles for the CG and CGC classes and their principal wavelength window. These parameters were estimated using *MatLab*[®] R2013a (The MathWorks Inc., Natick, USA) with a non-linear least squares fit algorithm. The procedure was performed in the 1000–2500-nm wavelength range.

Multivariate analysis methods

Principal component analysis

The HSI-NIR data set of 51 average samples of the CG and CGC strains were analysed with the principal component analysis (PCA) algorithm⁵⁵ using *The Unscrambler X.3* software (*CAMO*, Oslo, Norway).

Analysis of SPA-LDA and SIMCA

The execution of the SPA-LDA algorithm uses three sets of sample spectral matrices, referred to as training set and leave-one-out cross validation, and external prediction set, that were separated using the KS algorithm.⁵⁶ For this procedure, the samples from CG strains were separated from 3 and 2, and the samples from CGC strains were separated from 25 and 21 samples training plus cross validation sets and prediction sets, respectively. A PCA was run for the sample groups of the CG and CGC strains to combine with the PCR (rdDNA method) and KS algorithm results so that they used the same prediction

samples. Therefore, the same samples sets were used for SIMCA and SPA-LDA.

Index class with numbers 1 and 2 were assigned to the CG and CGC fungi samples, respectively. The numbers of variables selected by SPA-LDA are at least one and at most 20 variables or wavelengths. The SPA-LDA response by validation provides the error numbers of the test sets (cross validation) and the prediction set as well as the cost obtained by the test set with the optimal number of variables. For the identification of a sample and its class when there is yes or no for the occurrence of errors, the multi-LDA file (multilda.m) was used.

Box plot for SPA

The variables selected by the SPA in the HSI-NIR data were identified and processed in Microsoft Excel® worksheets to apply the box plot method for both samples of the CG and CGC strains. The lower limit (*LL*) and upper limit (*UL*) values were inserted into Equations 5 and 6 and represent the intensities of the boundaries that may be an outlier between fungi groups sample sets.

$$LL = q(0.25) - 1.5 [q(0.75) - q(0.25)] \quad (5)$$

$$UL = q(0.75) + 1.5 [q(0.75) - q(0.25)] \quad (6)$$

The occurrence of GC is less common than the CGC. For this reason, only five mean CG samples were obtained with HSI-NIR, which are statistically insufficient to perform the SPA algorithm with adequate validation. Therefore, CG and CGC samples means box plot analysis allows the NIR hyperspectra to be analysed and compared to the results obtained by the SPA model.

Results and discussion

Pathogenicity test

All five CGC strains tested induced ramulosis symptoms in the inoculated cotton plants at 36 DAE (Figure 3A–3E). Inoculated plants with CGC strains showed initial symptoms of ramulosis 10 days after inoculation (DAI), including star-shaped lesions (Figure 3A) and small lesions on leaves (Figure 3B); after 20–30 DAI, the ramulosis fungus caused death of the apical meristem (Figure 3C and 3D), which resulted in the shortening of internodes, stunting and the excessive development of branches exhibiting witches' broom symptoms (Figure 3E). Plants inoculated with all three GC strains exhibited

symptoms of anthracnose, such as small lesions along the veins, at 20 DAI (Figure 3F), that expand to large necrotic areas around the main leaf veins 30 DAI with the pathogen (Figure 3G). Plants inoculated with sterilised distilled water (controls) showed no symptoms of the diseases.

Reference assays by molecular DNA techniques

The IGS region of the rDNA of CG and CGC isolates (Table 1) showed a difference in size when sequenced. The IGS region from CG was approximately 300bp longer than that of CGC, which allows their differentiation by a simple PCR diagnosis with fragment amplifications of 432bp and 148bp for CG and CGC, respectively; details are shown in Figure 4.

The molecular diagnosis allows identification using the minimum amount of genetic material by amplification in a few hours and provides the detection of genetic markers of infectious fungi,⁵⁷ as is the case of CG and CGC fungi.

Hyperspectral image in the NIR

All CG and CGC strains evaluated were successfully distinguished by NIR hyperspectral images and by the pattern recognition techniques used. These results corroborate the reference assay using the IGS region of the rDNA of CG and CGC to distinguish CG and CGC strains by PCR diagnosis with fragment amplification (Figure 4) as well as pathogenicity tests performed with three strains of CG and five strains of CGC (Figure 3), whose results agree with the hypothesis that CG and CGC, although closely related, are distinct pathogens responsible for two different diseases in cotton.^{9,12}

Although controversy regarding no consistent reports of morphological and molecular markers to differentiate cotton anthracnose and ramulosis fungi exists,^{10,11} our study corroborates studies conducted by Silva-Mann *et al.*⁹ and Salustiano *et al.*,¹² who used molecular DNA techniques to distinguish successfully CG and CGC fungi. The technique proposed in this work, compared with molecular DNA techniques used to differentiate CG and CGC, is a rapid and low-cost method for accurate identification of these fungi and a non-destructive technique requiring minimal or no sample preparation.^{17,18}

HSI-NIR reflectance

Equations for calculating the optical properties³⁵ of CG and CGC fungi are given for the scattering coefficient

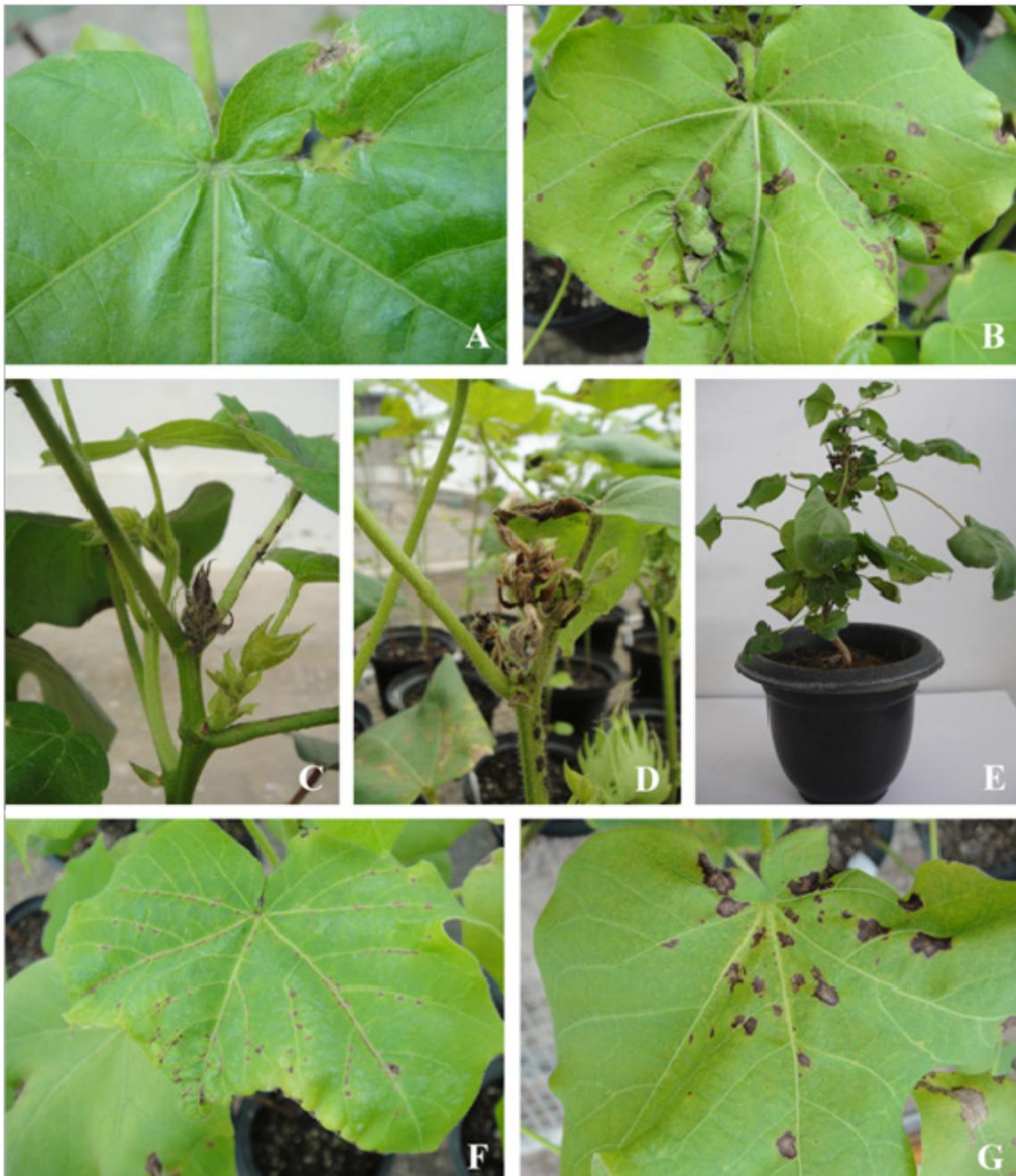


Figure 3. Symptoms of ramulosis and anthracnose in cotton plants inoculated during the pathogenicity test. A–B: star-shaped lesions and leaf lesions caused by *C. gossypii* var. *cephalosporioides* (CGC). C–D: death of the apical meristem. E: shortening of internodes, stunting and excessive development of branches (witches' broom symptoms). F–G: leaf anthracnose lesions caused by *C. gossypii* (CG).

cient, $\mu_s(\theta)$, by a power function, and for the absorption coefficient, $\mu_a(\theta)$, by means of a sigmoidal function. The parameters for μ_s and μ_a are organised in Table 2, and the results of the coefficients for CG and CGC are shown in the same spectral acquisition window (Figure 5).

The variance (V) defined by the equation:

$$V = 1 / (N - 1) \times \sum_{i=1}^N |A_i - \theta|^2,$$

has N numbers of optical coefficients that are of the same amplitude as the HSI-NIR spectrum. A_i represents the individual values of the optical coefficient and θ is the average of the optical coefficients. The variances of

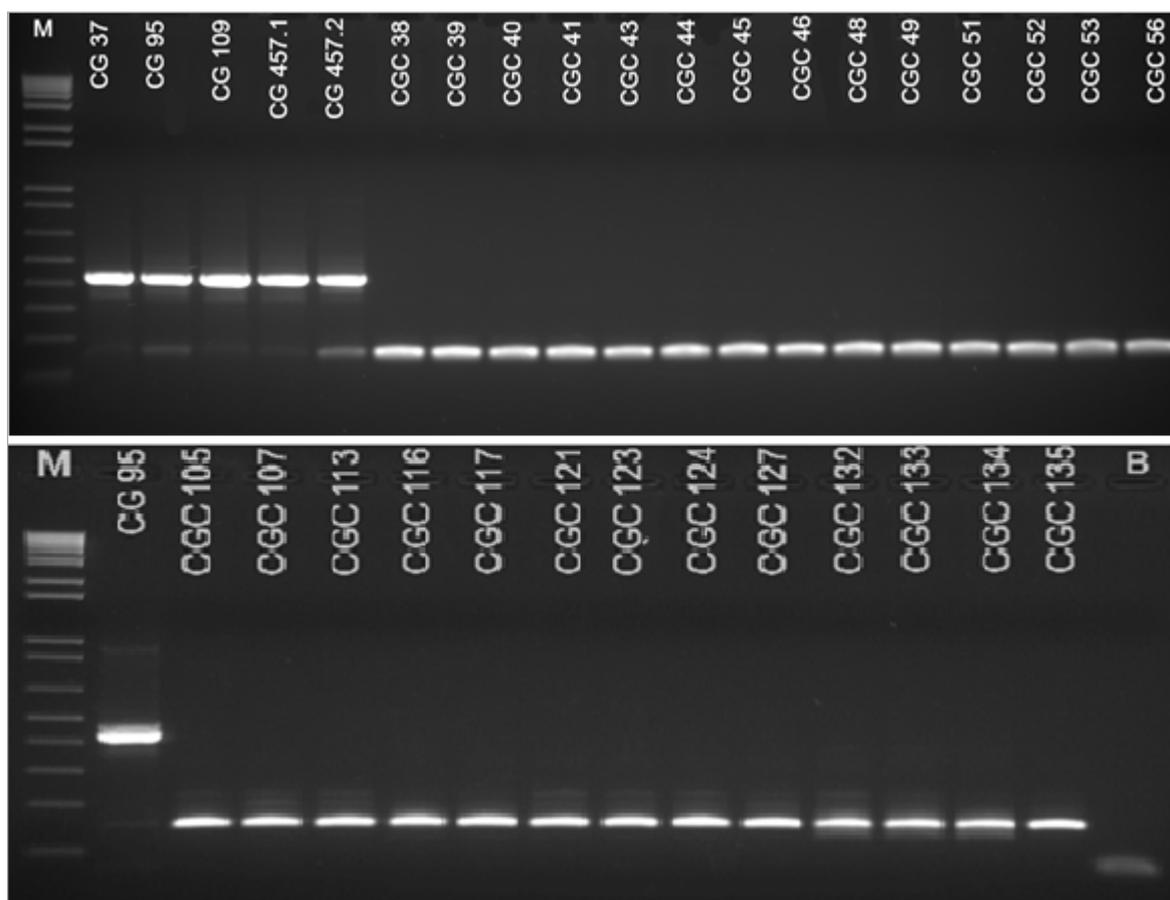


Figure 4. Differential molecular diagnosis using PCR amplification of the IGS region of different strains of CG and CGC. M represents the molecular marker 1 kb plus (Fermentas SM1331), and B is a negative control.

the scattering coefficients (μ_s) of CG and CGC strains are 1.37×10^{19} and 5.83×10^{-11} and are larger than the variances of the absorption coefficients (μ_a) (Table 2). The CG (Figure 5b) and CGC (Figure 5d) μ_a are of the same order of magnitude and showed that the light absorption signals are greater than the light scattering ones. The μ_s of CG (Figure 5a) is negative while for CGC (Figure 5c) it is positive, resulting in opposite signals. Besides that, i.e., the CG signals array has no light scattering while the CGC group does show scatter. Therefore, the variances of μ_s enable the determination of the CG and CGC classes.

The $\mu_s \ll \mu_a$ for the samples of CG (Figure 5a and 5b) and CGC strains (Figure 5c and 5d) evidences the absorption effect of the HSI-NIR spectra. The average μ_s of CG was less than that of the μ_s of CGC and the μ_a values were similar, although they were greater in amplitude. Therefore, the results of V and μ indicate the importance of not using the standard normal variate (SNV) in the pretreatment of the spectra, since CG and

CGC fungi have scattering and absorption coefficients that affect the separation of the CG and CGC classes. The non-use of techniques such as SNV is due to the geometric differences³³ that both fungi have. Therefore, SNV pretreatment removes the entire scattering effect due to morphological structures.⁵⁸ These effects are perfectly visualised in the HSI-NIR spectrum acquisition mode by reflectance.^{32,58}

Chemometric treatment of HSI-NIR spectra

The averages of the HSI-NIR spectra of each isolate sample were obtained. The profiles of the μ_s responded perfectly to CG and CGC classification in the HSI-NIR spectral range (1000–1381 nm) (Figure 5a and 5c) due to the presence of the water absorption signal from 1400 nm in the GC and CGC spectra (Figure 6).

In the PCA scores plot, samples of CG fungal strains clustered to the left of PC1, while CGC strains clustered to the right of PC1. PC1 explained 97.06% of the vari-

Table 2. Coefficients of the mathematical functions of scattering and absorption.

Scattering coefficient (μ_s): $\mu_s(\lambda) = a \times \lambda^b$ ($\lambda = \text{nm}$; $\mu_s = \text{mm}^{-1}$)				
Mean of isolates	A	B	Variance (V)	
CG	-0.827	-10.07	1.37×10^{19}	
CGC	6.50×10^{-5}	1.131	5.83×10^{-11}	
Absorption coefficient (μ_a): $\mu_a(\lambda) = \frac{a}{1 + \exp\left(-\frac{-x_0}{b}\right)}$ ($\lambda = \text{nm}$; $\mu_s = \text{mm}^{-1}$)				
Mean of isolates	A	b	x_0	Variance (V)
CG	0.2267	0.903	0.5489	4.44×10^{-5}
CGC	0.3061	0.903	0.5489	9.88×10^{-5}

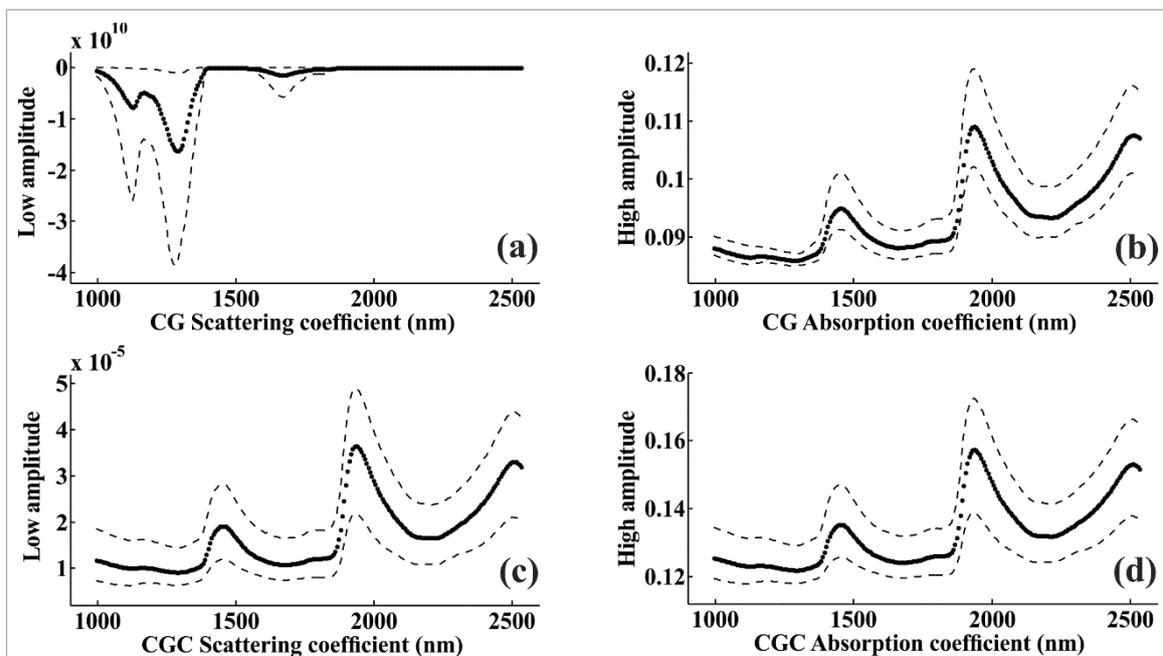


Figure 5. Scattering coefficient (μ_s) of CG (a) and CGC (c) and absorption coefficient (μ_a) of CG (b) and CGC (d). The μ_s values of CG are lower in amplitude than are the μ_s values of CGC. The dotted lines represent the minimum and maximum values of the mean (dotted) coefficients of the samples of the 5 CG and 46 CGC isolates.

ance; PC2, of 2.71%. The total explained variance of the fungi was 99.77%, but there was still some overlap between the CG and CGC strains because they were fungi of the same species, and the 1000–1381 nm spectral region is responsible for CG and CGC opposite signs of μ_s variance (Table 2a). After observing these differences of the spectra, the SIMCA model (Figure 7) was used to evaluate differences between the classes of samples of the CG and CGC strains, building a PCA model for each class with their respective figures of merit.

The HSI-NIR method with the data treated with the SIMCA and SPA-LDA algorithms is reliable and accurate (Figure 7). The SIMCA CG and CGC strains models were selected with a PC (both 83% and 96%, respectively), and the CG 457-2 and CGC 39 responses were not statistically significant (α) at the 5% level. The total data were equal to the genetic response of the PCR presenting two type II errors.

The SPA-LDA result for the 20 spectral variables occurred at a cost of 0.432, the lowest cost obtained

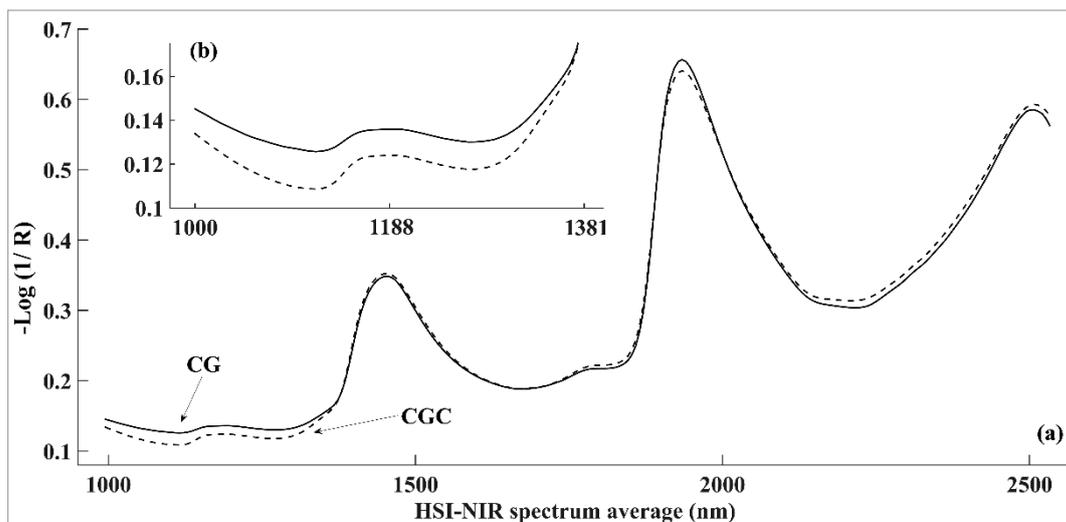


Figure 6. (a) Total HSI-NIR spectra of samples CG 37 and CGC 38 with the mean for all hyper-spectral images of the isolate and (b) the range of 1000–1381 nm that registered different values for the scattering reflectance (μ_s).

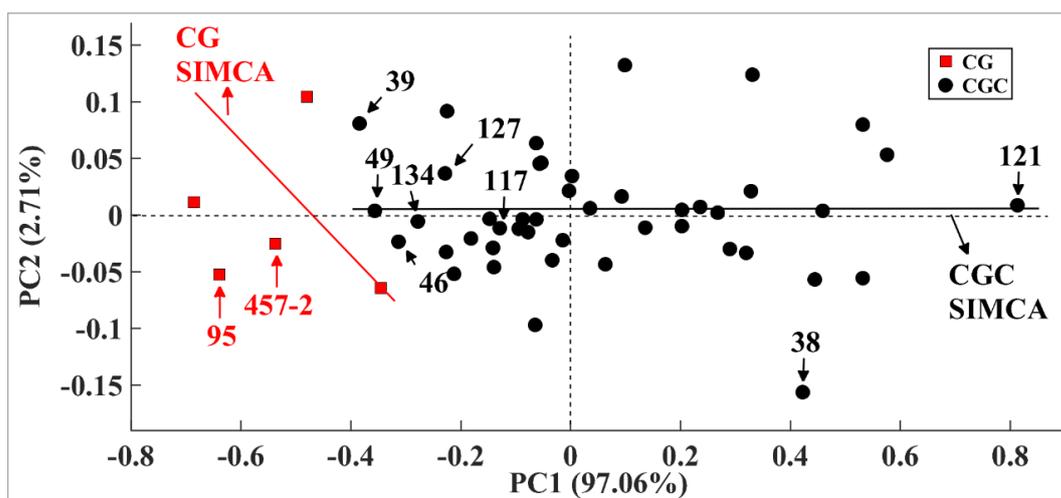


Figure 7. PC1 versus PC2 scores of the mean HSI-NIR spectra from 1000 nm to 1381 nm relative to the samples of the fungi CG (■) and CGC (●) with 99.77% of the variance explained. The centres of the SIMCA CG and CGC (linear segment) models are trained with samples determined the by PCR reference method (CGC 38, 46, 49, 121, 127 and 134) and show type II errors of classification (CG 457-2 and CGC 39).

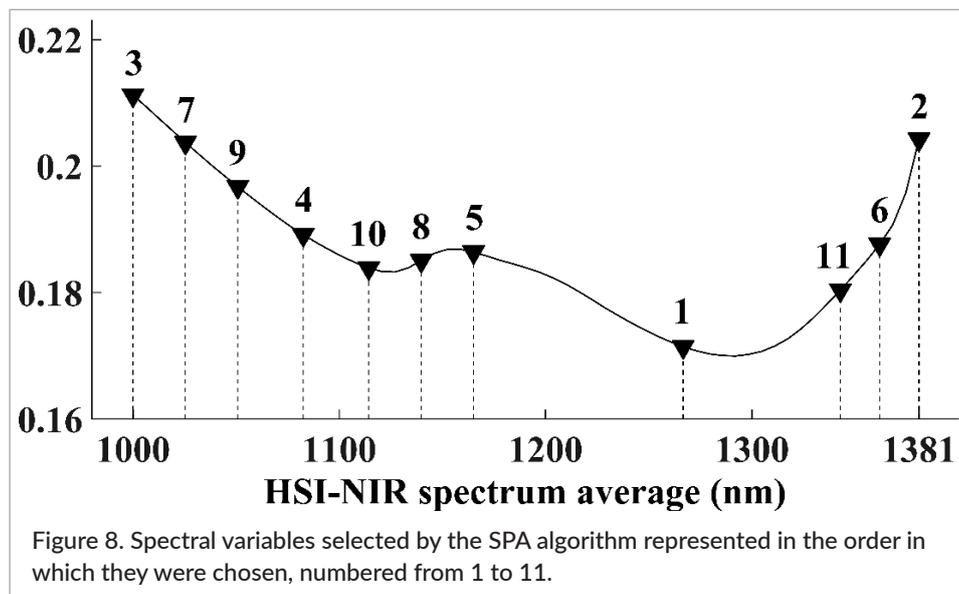
using 13 spectral variables, which indicated that sample CGC 117 was classified as an error. The CGC 117 sample is at the centre of the CGC class (Figure 7), showing that this error is due to a supersaturation in the SIMCA model for this class. Therefore, the optimal SPA-LDA result was calculated for 11 spectral variables at a cost of 0.447 and no error in the prediction set.

Among the 61 spectral variables of the HSI-NIR (1000–1381 nm) that were analysed and chosen by the PCA algorithm, 11 variables were chosen by the SPA algo-

riothm, which were the spectral variables (1) 1267, (2) 1381, (3) 1000, (4) 1082, (5) 1165, (6) 1362, (7) 1025, (8) 1139, (9) 1050, (10) 1114 and (11) 1342 nm (Figure 8).

An association of characteristic regions in the NIR can be interpreted based on known and already established bands in the literature for several types of compounds containing NH, CH and OH bonds.⁵⁹

The wavelengths of 1000–1114 nm, which correspond to the five variables selected by the SPA-LDA (Figure 8), correspond to the region of the second NH

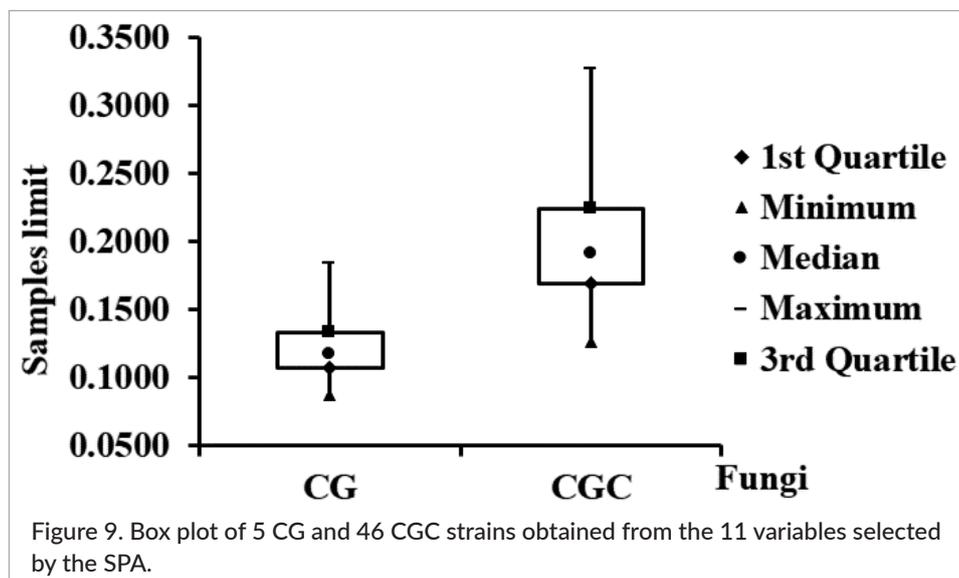


and CH overlap with the RNH₂ and ArCH groups and with the beginning of the CH₃ group. At 1139, 1165 and 1267 nm, the uptake of CH₃, CH₂ and CH groups occurs. Wavelengths 1342, 1362 and 1381 nm correspond to the first overlap regions of OH and CH with the chemical groups CH₃, CH₂ and ArOH. These are the chemical groups that enable information to differentiate the classes of diseases caused by GC and CGC.

The amount of CG fungal strains is very small compared to the amount of CGC fungal strains. However, a box plot was used to authenticate the calculations made by the SPA-LDA algorithm. The intensities of the 11 spectral variables selected resulted in 55 and 506 absorbance values for the 5 GC samples and the 46 CGC samples,

respectively, and the minimum, first quartile, median, third quartile and maximum were calculated. The amplitudes of the absorbance values obtained for the 11 variables chosen by the SPA presented differences between the first and third quartiles, which were 0.11 and 0.13 for the CG strains, respectively, and 0.17 and 0.22 for the CGC strains, respectively. Containing 50% of the different HSI-NIR variables between CG and CGC strains (Figure 9) promotes a perfect separation by the LDA method.

The minimum and maximum values of the spectral intensities of the CG strains were 0.09 and 0.18, respectively, and of the CGC strains were 0.13 and 0.33, respectively. The values adjacent to the *LL* and *UL* were 0.07 and 0.17 for CG, respectively, and 0.09 and 0.31



for CGC, respectively. Therefore, because the *LL* and *UL* values were lower than the minimum and maximum values of both CG and CGC fungi, there are no possible outliers for these two groups. Thus, the adjacent values are simply the points of the minimum and maximum data, which corroborates with the authenticity of the SPA-LDA model for the fungi HSI-NIR.

The region of 1000–1381 nm favours the absorption characteristics of the primary chemical compounds such as amino acids, enzymes and proteins. These compounds are directly linked to biochemical expression with respect to the structures of the genetic composition of the chemical groups CH₃, CH₂, CH, ArCH, ArOH and NH, which provide the intensities of the HSI-NIR.⁵⁹ Strategies allied to the SPA-LDA algorithm³⁰ result in 100% separation between the two classes of CG and CGC fungi.

Conclusion

In this work, a new analytical strategy was developed using HSI-NIR and data treatment strategies with the selection of variables such as SPA, construction of LDA and SIMCA classification models for samples of two species of cotton pathogenic fungi that are difficult to identify. The strategy used the scattering coefficients (μ_s) of HSI-NIR for absorption in the 1000–1381 nm range and required no spectral pretreatment. Therefore, in this study, we did not find type I errors at the 95% confidence level for CG and CGC fungal classification based on the selection of 11 spectral variables with the SPA-LDA model. The SIMCA produced only two type II errors, in which two CG 457-2 and CGC 39 prediction samples were classified into the two GC and CGC fungal classes. This result was confirmed using molecular markers and is attributed to the intraspecific and interspecific variation of the CG and CGC species by molecular descriptors. The combination of the HSI-NIR techniques with the SPA-LDA and SIMCA algorithms was efficient for the rapid, automated and non-destructive identification and classification of multiple CG and CGC fungi strains growing on PDA medium.

Acknowledgements

The authors would like to thank Embrapa – Empresa Brasileira de Pesquisa Agropecuária and IMAmt – Instituto

Matogrossense do Algodão (N. proc: 03.15.00.051.00.00). The authors also thank the Brazilian agencies CNPq and CAPES (15/ 2014) for scholarships supporting this work.

References

1. J.A. Bailey, C. Nash, L.W. Morgan, R.J. O'Connell and D.O. TeBeest, "Molecular taxonomy of colletotrichum species causing anthracnose on the malvaceae", *Phytopathology* **86**, 1076–1083 (1996). doi: <https://doi.org/10.1094/Phyto-86-1076>
2. A.S. Costa and C.G. Fraga-Júnior, "Sobre a natureza da ramulose ou superbrotoamento do algodoeiro", *J. Agron.* **2**, 151 (1939).
3. J. Abrahão and A.S. Costa, "Instruções para o reconhecimento da ramulose do algodoeiro", *O Biológico, São Paulo* **15**, 59 (1949).
4. E.A. Southworth, "Anthracnose of cotton", *J. Mycol.* **6**, 100–105 (1891). doi: <https://doi.org/10.2307/3752564>
5. G.F. Atkinson, "Anthracnose of cotton", *J. Mycol.* **6**, 173–184 (1891). doi: <https://doi.org/10.2307/3752660>
6. M.A.S. Tanaka, J.O.M. Menten and J.C. Machado, "Growth habit of *Colletotrichum gossypii* and *C. gossypii* var. *cephalosporioides* on cotton seeds", *Bragantia* **55(1)**, 95 (1996). doi: <https://doi.org/10.1590/S0006-87051996000100011>
7. T. Kirkpatrick and C. Rothrock, *Compendium of Cotton Diseases*, 2nd Edn. APS Press, St Paul, USA (2001).
8. I.A. Kroetz, "Portaria no. 47, de 26 de fevereiro de 2009", *Ministério da Agricultura, Pecuária e Abastecimento*. Available online: http://www.apps.agr.br/upload/ax4_0303200929901500_anexo-port-47-26-fev-2009.pdf [Accessed: 20 June 2017].
9. R. Silva-Mann, M.G.G.C. Vieira, J.C. Machado, J.R. Bernardino-Filho, K.C.C. Salgado and M.R. Stevens, "AFLP markers differentiate isolates of *Colletotrichum gossypii* from *C. gossypii* var. *cephalosporioides*", *Fitopatol. Bras.* **30(2)**, 169 (2005). doi: <https://doi.org/10.1590/S0100-41582005000200011>
10. K.D. Hyde, L. Cai, P.F. Cannon, J.A. Crouch, P.W. Crous, U. Damm, P.H. Goodwin, H. Chen, P.R. Johnston, E. Jones, Z.Y. Liu, E. Mckenzie, J. Moriwaki, P. Noireung, S.R. Pennycook,

- L.H. Pfenning, H. Prihastuti, T. Sato, R.G. Shivas, Y.P. Tan, P. Taylor, B.S. Weir, Y.L. Yang and J.Z. Zhang, "Colletotrichum – names in current use", *Fungal Divers.* **39**, 147 (2009).
11. B.C. Sutton, "The genus *Glomerella* and its anamorph *Colletotrichum*", in *Colletotrichum: Biology, Pathology and Control*, Ed by J.A. Bailey and M.J. Jeger. CAB International, Wallingford, UK, p. 1 (1992).
 12. M.E. Salustiano, M.N. Rondon, L.M. Abreu, S.S. Costa, J.C. Machado and L.H. Pfenning, "The etiological agent of cotton ramulosis represents a single phylogenetic lineage within the *Colletotrichum gloeosporioides* species complex", *Trop. Plant Pathol.* **39(5)**, 357 (2014). doi: <https://doi.org/10.1590/S1982-56762014000500002>
 13. B.S. Weir, P.R. Johnston and U. Damm, "The *Colletotrichum gloeosporioides* species complex", *Stud. Mycol.* **73**, 115–180 (2012). doi: <https://doi.org/10.3114/sim0011>
 14. G. ElMasry and D.-W. Sun, "Chapter 1 – principles of hyperspectral imaging technology", in *Hyperspectral Imaging Food Qual. Anal. Control*, Ed by D.-W. Sun. Academic Press, pp. 3–43 (2010). doi: <https://doi.org/10.1016/B978-0-12-374753-2.10001-2>
 15. A.A. Gowen, Y. Feng, E. Gaston and V. Valdramidis, "Recent applications of hyperspectral imaging in microbiology", *Talanta* **137**, 43–54 (2015). doi: <https://doi.org/10.1016/j.talanta.2015.01.012>
 16. A. Ziadi, X. Maldague, L. Saucier, C. Duchesne and R. Gosselin, "Visible and near-infrared light transmission: a hybrid imaging method for non-destructive meat quality evaluation", *Infrared Phys. Technol.* **55(5)**, 412–420 (2012). doi: <https://doi.org/10.1016/j.infrared.2012.05.004>
 17. M.T. Islam, N. Scoutaris, M. Maniruzzaman, H.G. Moradiya, S.A. Halsey, M.S.A. Bradley, B.Z. Chowdhry, M.J. Snowden and D. Douroumis, "Implementation of transmission NIR as a PAT tool for monitoring drug transformation during HME processing", *Eur. J. Pharm. Biopharm.* **96**, 106–116 (2015). doi: <https://doi.org/10.1016/j.ejpb.2015.06.021>
 18. S.H. Gordon, R.W. Jones, J.F. McClelland, D.T. Wicklow and R.V. Greene, "Transient infrared spectroscopy for detection of toxigenic fungi in corn: potential for on-line evaluation", *J. Agric. Food Chem.* **47(12)**, 5267–5272 (1999). doi: <https://doi.org/10.1021/jf990011f>
 19. F.E. Dowell, M.S. Ram and L.M. Seitz, "Predicting scab, vomitoxin, and ergosterol in single wheat kernels using near-infrared spectroscopy", *Cereal Chem. J.* **76(4)**, 573–576 (1999). doi: <https://doi.org/10.1094/CCHEM.1999.76.4.573>
 20. H. Pettersson and L. Aberg, "Near infrared spectroscopy for determination of mycotoxins in cereals", *Food Control* **14(4)**, 229–232 (2003). doi: [https://doi.org/10.1016/S0956-7135\(03\)00011-2](https://doi.org/10.1016/S0956-7135(03)00011-2)
 21. V. Fernández-Ibañez, A. Soldado, A. Martínez-Fernández and B. Roza-Delgado, "Application of near infrared spectroscopy for rapid detection of aflatoxin B1 in maize and barley as analytical quality assessment", *Food Chem.* **113(2)**, 629–634 (2009). doi: <https://doi.org/10.1016/j.food-chem.2008.07.049>
 22. N. Berardo, V. Pisacane, P. Battilani, A. Scandolaro, A. Pietri and A. Marocco, "Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy", *J. Agric. Food Chem.* **53(21)**, 8128–8134 (2005). doi: <https://doi.org/10.1021/jf0512297>
 23. C. Singh, D. Jayas, J. Paliwal and N. White, "Fungal damage detection in wheat using short-wave near-infrared hyperspectral and digital colour imaging", *Int. J. Food Prop.* **15(1)**, 11–24 (2012). doi: <https://doi.org/10.1080/10942911003687223>
 24. J. Qin, T. Burks, M. Ritenour and W. Bonn, "Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence", *J. Food Eng.* **93(2)**, 183–191 (2009). <https://doi.org/10.1016/j.jfoodeng.2009.01.014>
 25. J.M. Amigo, H. Babamoradi and S. Elcoroaristizabal, "Hyperspectral image analysis. A tutorial", *Anal. Chim. Acta* **896**, 34–51 (2015). doi: <https://doi.org/10.1016/j.aca.2015.09.030>
 26. C. Xie, C. Yang and Y. He, "Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities", *Comput. Electron. Agric.* **135**, 154–162 (2017). doi: <https://doi.org/10.1016/j.compag.2016.12.015>
 27. J. Riedl, S. Esslinger and C. Fauhl-Hassek, "Review of validation and reporting of non-targeted fingerprinting approaches for food authentication", *Anal. Chim. Acta* **885**, 17–32 (2015). doi: <https://doi.org/10.1016/j.aca.2015.06.003>
 28. S.F.C. Soares, E.P. Medeiros, C. Pasquini, C.L. Morello, R.K.H. Galvão and M.C.U. Araujo,

- “Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging”, *Anal. Methods* **8(48)**, 8498–8505 (2016). doi: <https://doi.org/10.1039/C6AY02896A>
29. R. Calvini, J.M. Amigo and A. Ulrici, “Transferring results from NIR-hyperspectral to NIR-multispectral imaging systems: a filter-based simulation applied to the classification of arabica and robusta green coffee”, *Anal. Chim. Acta* **967**, 33–41 (2017). doi: <https://doi.org/10.1016/j.aca.2017.03.011>
 30. M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, P.N.T. Moreira, O.D.P. Neto, G.E. José and T.C.B. Saldanha, “The successive projections algorithm for spectral variable selection in classification problems”, *Chemometr. Intell. Lab. Syst.* **78(1)**, 11–18 (2005). doi: <http://dx.doi.org/10.1016/j.chemo-lab.2004.12.001>
 31. D. Wu and D. Sun, “Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals”, *Innov. Food Sci. Emerg. Technol.* **19**, 1–14 (2013). doi: <https://doi.org/10.1016/j.ifset.2013.04.014>
 32. J. Qin, “Hyperspectral imaging instruments”, in *Hyperspectral Imaging for Food Quality Analysis and Control*, Elsevier, Beltsville, Maryland, USA, pp. 129–172 (2010). doi: <https://doi.org/10.1016/B978-0-12-374753-2.10005-X>
 33. H.F. Grahn and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*. John Wiley & Sons, Chichester, UK (2007). doi: <https://doi.org/10.1002/9780470010884>
 34. T.J. Farrell, M.S. Patterson and B. Wilson, “A diffusion model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties *in vivo*”, *Med. Phys.* **19(4)**, 879–888 (1992). doi: <https://doi.org/10.1118/1.596777>
 35. R. Michels, F. Foschum and A. Kienle, “Optical properties of fat emulsions”, *Opt. Express* **16(8)**, 5907–5925 (2008). doi: <https://doi.org/10.1364/OE.16.005907>
 36. X. Cheng and D.A. Boas, “Diffuse optical reflection tomography with continuous-wave illumination”, *Opt. Express* **3(3)**, 118–123 (1998). doi: <https://doi.org/10.1364/OE.3.000118>
 37. N.A. Carbone, G.R. Baez, H.A. García, M.V. Waks Serra, H.O. Di Rocco, D.I. Iriarte, J.A. Pomarico, D. Grosenick and R. Macdonald, “Diffuse reflectance optical topography: location of inclusions in 3D and detectability limits”, *Biomed. Opt. Express* **5(5)**, 1336–1354 (2014). doi: <https://doi.org/10.1364/BOE.5.001336>
 38. D.A. Boas, L.E. Campbell and A.G. Yodh, “Scattering and imaging with diffusing temporal field correlations”, *Phys. Rev. Lett.* **75(9)**, 1855 (1995). doi: <https://doi.org/10.1103/PhysRevLett.75.1855>
 39. A. Yodh and B. Chance, “Spectroscopy and imaging with diffusing light”, *Phys. Today* **48(3)**, 34 (1995). doi: <https://doi.org/10.1063/1.881445>
 40. A.P. Gibson, J.C. Hebden and S.R. Arridge, “Recent advances in diffuse optical imaging”, *Phys. Med. Biol.* **50(4)**, R1 (2005). doi: <https://doi.org/10.1088/0031-9155/50/4/R01>
 41. A.H. Hielscher, A.Y. Bluestone, G.S. Abdoulaev, A.D. Klose, J. Lasker, M. Stewart, U. Netz and J. Beuthan, “Near-infrared diffuse optical tomography”, *Dis. Markers* **18(5–6)**, 313–337 (2002). doi: <https://doi.org/10.1155/2002/164252>
 42. D.A. Boas, M.A. O’Leary, B. Chance and A.G. Yodh, “Detection and characterization of optical inhomogeneities with diffuse photon density waves: a signal-to-noise analysis”, *Appl. Opt.* **36(1)**, 75–92 (1997). doi: <https://doi.org/10.1364/AO.36.000075>
 43. M.C.W. van Rossum and T.M. Nieuwenhuizen, “Multiple scattering of classical waves: microscopy, mesoscopy, and diffusion”, *Rev. Mod. Phys.* **71(1)**, 313 (1999). doi: <https://doi.org/10.1103/RevModPhys.71.313>
 44. T. Durduran, A.G. Yodh, B. Chance and D.A. Boas, “Does the photon-diffusion coefficient depend on absorption?”, *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* **14(12)**, 3358–3365 (1997). doi: <https://doi.org/10.1364/JOSAA.14.003358>
 45. E. Sutherland, S.M. Mereer, M. Everist and D.G. Leaist, “Diffusion in solutions of micelles. What does dynamic light scattering measure?”, *J. Chem. Eng. Data* **54(2)**, 272–278 (2009). doi: <https://doi.org/10.1021/je800284g>
 46. R. Elaloufi, R. Carminati and J.-J. Greffet, “Definition of the diffusion coefficient in scattering and absorbing media”, *J. Opt. Soc. Am. A* **20(4)**, 678–685 (2003). doi: <https://doi.org/10.1364/JOSAA.20.000678>
 47. Y.H. Kim, H.S. Jang and J.Y. Jeon, “Characterizing diffusive surfaces using scattering and diffusion coefficients”, *Appl. Acoust.* **72(11)**, 899–905 (2011). doi: <https://doi.org/10.1016/j.apacoust.2011.05.006>

48. J. Stam, "Multiple scattering as a diffusion process", *Proc. Eurographics Work. Render. Tech.* 41–50 (1995). doi: https://doi.org/10.1007/978-3-7091-9430-0_5
49. C. Culbertson, "Diffusion coefficient measurements in microfluidic devices", *Talanta* **56(2)**, 365–373 (2002). doi: [https://doi.org/10.1016/S0039-9140\(01\)00602-6](https://doi.org/10.1016/S0039-9140(01)00602-6)
50. A.D. McBrady, R. Chantiwas, A.K. Torgerson, K. Grudpan and R.E. Synovec, "An absorbance-based micro-fluidic sensor for diffusion coefficient and molar mass determinations", *Anal. Chim. Acta* **575(2)**, 151–158 (2006). doi: <https://doi.org/10.1016/j.aca.2006.05.083>
51. R. Auria, J. Palacios and S. Revah, "Determination of the interparticular effective diffusion coefficient for CO₂ and O₂ in solid state fermentation", *Biotechnol. Bioeng.* **39(9)**, 898–902 (1992). doi: <https://doi.org/10.1002/bit.260390903>
52. L.S. Zhang, C.P. Zhang, X.Y. Wang, S.W. Qi, T. Xu, J.G. Tian and G.Y. Zhang, "Influence of boundary condition and diffusion coefficient on the accuracy of diffusion theory in steady-state spatially resolved diffuse reflectance of biological tissues", *Chinese Phys. Lett.* **19(12)**, 1874 (2002). doi: <https://doi.org/10.1088/0256-307X/19/12/340>
53. J. Doyle, "DNA protocols for plants", in *Molecular Techniques in Taxonomy*, Ed by G.M. Hewitt, A.W.B. Johnston and J.P.W. Young. NATO ASI Series (Series H: Cell Biology), Vol. 57, Ch. 18, pp. 283–293, Springer, Berlin, Heidelberg (1991). doi: https://doi.org/10.1007/978-3-642-83962-7_18
54. A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data by simplified least square procedures", *Anal. Chem.* **36(8)**, 1627–1639 (1964). doi: <https://doi.org/10.1021/ac60214a047>
55. R.G. Brereton (Ed.), *Multivariate Pattern Recognition in Chemometrics, Volume 9*. Elsevier Science, New York (1992).
56. R.W. Kennard and L.A. Stone, "Computer aided design of experiments", *Technometrics* **11(1)**, 137–148 (1969). doi: <https://doi.org/10.1080/00401706.1969.10490666>
57. K. Mullis, F. Faloon, S. Scharf, R. Saiki, G. Horn and H. Erlich, "Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction", *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263–273 (1986). doi: <https://doi.org/10.1101/SQB.1986.051.01.032>
58. A.M. Rady, D.E. Guyer, W. Kirk and I.R. Donis-González, "The potential use of visible/near infrared spectroscopy and hyperspectral imaging to predict processing-related constituents of potatoes", *J. Food Eng.* **135**, 11–25 (2014). doi: <https://doi.org/10.1016/j.jfoodeng.2014.02.021>
59. D.A. Burns and E.W. Ciurczak, *Handbook of Near-Infrared Analysis*, 3rd Edn. CRC Press (2007).