

applied spectroscopy

Discrimination of Corn from Monocotyledonous Weeds with Ultraviolet (UV) Induced Fluorescence

BERNARD PANNETON,* SERGE GUILLAUME, GUY SAMSON, and JEAN-MICHEL ROGER

Horticultural R&D Centre, Agriculture and Agri-Food Canada, St-Jean-sur-Richelieu, Qc, Canada J3B 3E6 (B.P.); Cemagref, UMR ITAP, 34196 Montpellier, France (S.G., J.-M.R.); and Département de chimie-biologie, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, G9A 5H7 (G.S.)

Discrimination of Corn from Monocotyledonous Weeds with Ultraviolet (UV) Induced Fluorescence

BERNARD PANNETON,* SERGE GUILLAUME, GUY SAMSON, and JEAN-MICHEL ROGER

Horticultural R&D Centre, Agriculture and Agri-Food Canada, St-Jean-sur-Richelieu, Qc, Canada J3B 3E6 (B.P.); Cemagref, UMR ITAP, 34196 Montpellier, France (S.G., J.-M.R.); and Département de chimie-biologie, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, G9A 5H7 (G.S.)

In production agriculture, savings in herbicides can be achieved if weeds can be discriminated from crop, allowing the targeting of weed control to weed-infested areas only. Previous studies demonstrated the potential of ultraviolet (UV) induced fluorescence to discriminate corn from weeds and recently, robust models have been obtained for the discrimination between monocots (including corn) and dicots. Here, we developed a new approach to achieve robust discrimination of monocot weeds from corn. To this end, four corn hybrids (Elite 60T05, Monsanto DKC 26-78, Pioneer 39Y85 (RR), and Syngenta N2555 (Bt, LL)) and four monocot weeds (*Digitaria ischaemum* (Schreb.) I, *Echinochloa crus-galli* (L.) Beauv., *Panicum capillare* (L.), and *Setaria glauca* (L.) Beauv.) were grown either in a greenhouse or in a growth cabinet and UV (327 nm) induced fluorescence spectra (400 to 755 nm) were measured under controlled or uncontrolled ambient light intensity and temperature. This resulted in three contrasting data sets suitable for testing the robustness of discrimination models. In the blue-green region (400 to 550 nm), the shape of the spectra did not contain any useful information for discrimination. Therefore, the integral of the blue-green region (415 to 455 nm) was used as a normalizing factor for the red fluorescence intensity (670 to 755 nm). The shape of the normalized red fluorescence spectra did not contribute to the discrimination and in the end, only the integral of the normalized red fluorescence intensity was left as a single discriminant variable. Applying a threshold on this variable minimizing the classification error resulted in calibration errors ranging from 14.2% to 15.8%, but this threshold varied largely between data sets. Therefore, to achieve robustness, a model calibration scheme was developed based on the collection of a calibration data set from 75 corn plants. From this set, a new threshold can be estimated as the 85% quantile on the cumulative frequency curve of the integral of the normalized red fluorescence. With this approach the classification error was nearly constant (16.0% to 18.5%), thereby indicating the potential of UV-induced fluorescence to reliably discriminate corn from monocot weeds.

Index Headings: Fluorescence; Weeds; Corn; Monocots; Discrimination; Model robustness; Site-specific weed control.

INTRODUCTION

Sensor-based automatic discrimination of weeds from a crop could be of great benefit for production agriculture. With such a sensing system, field areas infested by weeds can be identified and control measures can be applied only to these areas. In all cases, limiting application of herbicides to a fraction of the cultivated fields results in time and cost savings for the farmer and lowered environmental impacts.

Previous studies indicated that UV-induced plant fluorescence can be used to discriminate plant groups.¹ Under UV excitation, plants can emit a blue-green fluorescence (BGF)

with a wide peak around 440 nm and also the chlorophyll fluorescence (ChlF), characterized by its two peaks in the red and far-red regions (685 and 735 nm) of the spectrum. The characteristics of the fluorescence emission spectra depend on different leaf properties, notably the concentrations of ferulic acid (the main emitter of BGF) and chlorophylls, and also the presence of non-fluorescent UV-absorbing compounds in leaf epidermis that decrease the UV excitation of chlorophylls in the leaf mesophyll.^{2,3} For plants grown under similar conditions and having similar developmental stage, these leaf properties vary according to plant species. Therefore, the characteristics of plant emission fluorescence spectra represent distinct signatures that may be used for plant discrimination. In the context of plant discrimination, several factors can affect the fluorescence signal and these were reviewed in a previous paper.⁴

The potential of UV-induced plant fluorescence spectra for discriminating between plant groups or plant species has been evaluated. Early work¹ demonstrated that dicotyledonous plants (dicots) can be distinguished from herbaceous monocotyledonous plants (monocots) based on the ratio F_{440}/F_{685} (F_{440} is the fluorescence intensity measured at 440 nm). The content of ferulic acid in monocotyledonous plants, and particularly in species of the *Poaceae* family, is several times higher than in dicotyledonous plants. In consequence, BGF emission is more intense in monocot leaves than in dicot leaves, resulting in higher F_{440}/F_{685} ratio.⁵ Also, the ratio F_{685}/F_{735} was used to discriminate four plant species: peas, barley, clover, and Shepherd's purse.⁶ But since this ratio is mainly influenced by leaf chlorophyll-*a* concentrations,⁷ the robustness of this fluorescence ratio for plant discrimination may be limited.

Recently, Panneton and co-workers specifically tested the potential of UV-induced fluorescence spectra to discriminate weeds from crop by using a larger number of species that are relevant to a crop-weed field environment. They developed a model calibrated for four corn hybrids, four monocot weeds, and four dicot weeds and obtained a cross-validation error of 8.2%.⁸ It is noteworthy that this low calibration error was obtained with a data set composed of fluorescence spectra measured on leaves from plants of different ages (10 to 30 days after emergence) and measured from different positions (leaf base and leaf apex), two factors known to significantly affect intrinsic leaf properties and thereby plant fluorescence emission spectrum.⁹ On the same group of plant species, it was shown that by using proper normalization¹⁰ robust discrimination between monocots (including corn) and dicots can be performed with a classification error in prediction between 1.5% and 5.2%.⁴ This was achieved using the average normalized signal in two bands: $F_{400-425}$ and $F_{425-490}$. Regarding the discrimination of monocot weeds from corn,

Received 16 August 2010; accepted 4 October 2010.

* Author to whom correspondence should be sent. E-mail: Bernard.panneton@agr.gc.ca.

DOI: 10.1366/10-06100

TABLE I. List of weed species and corn hybrids.

Plant group	Species or corn hybrids
Corn hybrids	Elite 60T05
	Monsanto DKC 26-78
	Pioneer 39Y85 (RR)
	Syngenta N2555 (Bt, LL)
Monocotyledonous weeds	<i>Digitaria ischaemum</i> (Schreb.) I
	<i>Echinochloa crus-galli</i> (L.) Beauv.
	<i>Panicum capillare</i> (L.)
	<i>Setaria glauca</i> (L.) Beauv.

partial least squares discriminant analyses (PLSDA¹¹) based on either the full normalized spectra (400 to 755 nm) or the blue-green portion (400 to 625 nm) resulted in calibration errors ranging from 4.8% to 13.6%. However in this case, the prediction errors ranged from 11% up to 50%, indicating a clear lack of robustness⁴ that impedes so far the application of UV-induced fluorescence for weed-corn discrimination in the field.

In this context, our main objective was to develop an alternative analytical approach that could improve the accuracy and the robustness of UV-induced fluorescence to discriminate monocot weeds and corn. To test the robustness of our models, we collected three data sets of plants grown under different conditions (greenhouse or growth chamber) and/or whose UV-induced fluorescence was measured under different ambient conditions (controlled or uncontrolled light intensity and air temperature). Also, to introduce intra-set variability, fluorescence was measured from different leaf positions and leaf ages. The results showed that the inter-set variability had the larger impacts on the robustness of our models. To minimize these effects, we proposed a straightforward recalibration approach that sets a threshold value resulting in nearly constant classification errors (16.0% to 18.5%). This approach could be considered as a second step following a first discrimination of dicots from monocots (including corn).

MATERIALS AND METHODS

The experiments spanned over three years (2005, 2006, and 2007) and all spectra from a single year will be referred to as a data set. Each year, the same species were used. These were four corn (*Zea mays* L.) hybrids and four annual monocotyledonous grass species (Table I). Corn is a monocotyledonous plant. In 2005 and 2006, plants were grown in a growth chamber and in a greenhouse in 2007 (Table II). In 2005 and 2006, light was provided by 1000 W metal halide lamps at a distance of approximately one meter from the pots, giving about 500 $\mu\text{mol}/\text{m}^2/\text{s}$. In the greenhouse, high-pressure sodium lamps were used to supplement sunlight. For all years, the photoperiod was 16 hours of light and 8 hours of darkness. In the growth chamber, the temperature was set to 20 °C during the day and 12 °C at night, with a plateau of one hour at 16 °C between each change of temperature. In the greenhouse, the maximum temperature was set to 24 °C and night temperature was maintained above 12 °C. Plants were grown in 1.07 L pots (127 mm dia.) of soil-less mix (Promix BX, Premier Horticulture, Quebec, Canada). Nutrients (N-P-K rating: 20-12-20 at 95 g/L) were dissolved in tap water and were applied at every watering. Care was taken to avoid systematic temperature and lighting gradients by moving pots around every other day.

TABLE II. Growth and measurement environments and sample sizes for each year.

	2005	2006	2007
Growth environment	Chamber	Chamber	Greenhouse
Measurement environment	Greenhouse	Chamber	Greenhouse
Sample size, corn	375	128	121
Sample size, monocotyledonous weeds	362	177	102

After sowing, the date of emergence was recorded for each pot. To introduce variations in the leaf fluorescence emission spectra within each data set and hence test the robustness of the models, fluorescence measurements were made between 15 and 20 days after emergence (stage 1) and again on the same plants, between 25 and 30 days after emergence (stage 2). At both stages, two measurements were performed on the uppermost fully developed leaf: one measurement on a point near the base of the leaf (lower 25% of the leaf blade) and another one near its apex (top 25% of the leaf blade).

In 2005 and 2007, measurements were performed in a greenhouse under natural daylight (Table II). Plants were placed in the greenhouse one hour before measurements. In 2006, measurements were performed under metal halide lamps at 500 $\mu\text{mol}/\text{m}^2/\text{s}$ and 20 °C. In 2006, the ambient conditions were stable and selected to be close to the mean conditions obtained in 2005 and 2007 in the greenhouse environment (averages of 440 $\mu\text{mol}/\text{m}^2/\text{s}$ and 22 °C). Each year, the experiment was repeated in time on three occasions in 2005 and 2006 and, on two occasions in 2007 (3 or 2 blocks of data). In each block, there were 8, 4, and 4 specimens of each species/hybrid for 2005, 2006, and 2007, respectively. Therefore, the experiments were planned to provide a total of 1408 spectra (3 data sets, 2 or 3 blocks, 2 plant groups, 4 hybrid/species per group, 4 or 8 replicates per hybrids/species, and 4 readings for each). In the end, some data were rejected for various reasons (growth problems, instrumentation problems, etc.) yielding a validated data set composed of 1265 spectra. The sample sizes for each of the three plant groups are given in Table II. The monocotyledonous weeds will be referred to as monocots.

FLUORESCENCE MEASUREMENTS

As illustrated in Fig. 1, plant fluorescence was induced by a xenon flash lamp (Spectra-physics Series Q Housing 60000 with a 5 J Xe pulsed arc lamp, Newport Corporation-Oriel Products, Stratford, CT) controlled by an Oriel 68826 power supply giving a 9 μs pulse width (Newport Corporation-Oriel Products, Stratford, CT). The flash output was coupled to a high-grade fused-silica fiber-optic bundle (3.2 mm diameter, Oriel 77578) using a condensing lens assembly (Oriel 60076) and a bandpass filter centered at 327 nm (20 nm full width at half-maximum (FWHM)). The induced fluorescence was collected by another fiber-optic bundle (high-grade fused-silica, SMA to 200 $\mu\text{m} \times 6$ mm slit, 0.22 NA, Oriel 77532) and transported to the spectrograph (Oriel MS125 1/8m). Using a length gauge, both fiber optics were positioned 5 mm above the leaf and pointed to the same spot on the leaf, 2.3 mm in diameter. The leaf blade was positioned perpendicular to the probe as judged by the operator. The spectrograph was modified by the insertion of a high-pass filter at the input port (400 nm: 5% at 388 nm and 80% at 405 nm) to cut off second-order effects. An intensified charge-coupled device (ICCD)

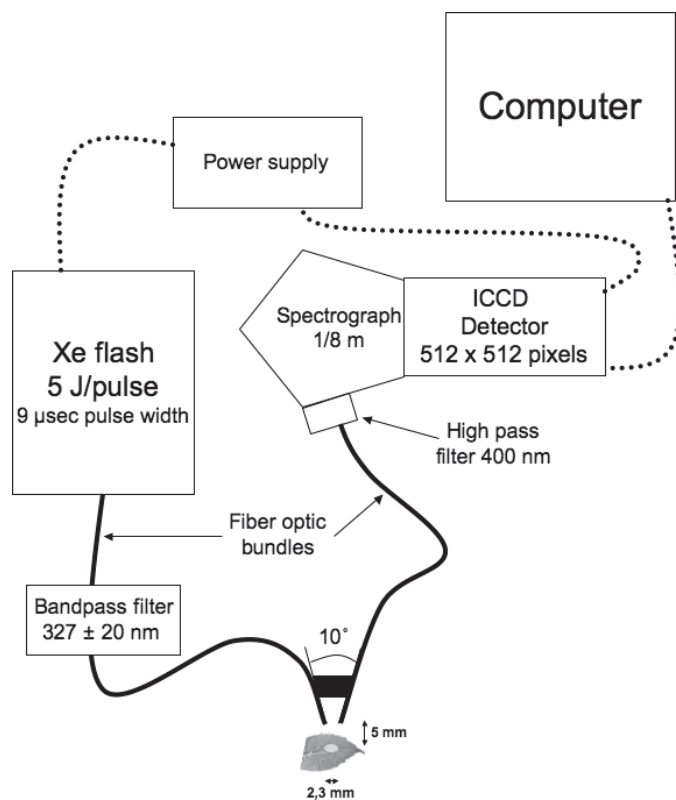


FIG. 1. Schematic representation of the fluorometer.

detector (Andor, DH 712-18F/03, 5 ns, Phosphor P43, Andor Technology PLC, Belfast, Northern Ireland) was coupled to the spectrograph to record the spectrum in the range from 398 to 760 nm (378 pixels or wavebands).

Fluorescence signals were acquired using the same technique as that of Norikane and Kuruta.¹² Under ambient light (sunlight in 2005 and 2007 or lamps in 2006) and without the UV excitation, 11 background spectra were acquired at 10 Hz and averaged. Then, 11 raw spectra were acquired under UV flash excitation (10 Hz) and averaged. The difference between the two resulting spectra is the pure induced fluorescence signal. The raw fluorescence spectra were smoothed with a moving average filter covering 15 channels (14.4 nm). Measurement repeatability was assessed by plotting the mean and standard deviation from the mean of the 11 raw spectra after background subtraction (Fig. 2). For quality control, means of all the spectra acquired in a given year were plotted together. This plot revealed a slight shift in wavelength for the 2005 data set (mean 2005 spectrum shifted by 3 to 4 nm towards higher wavelength). The most likely cause for this shift is a misalignment of the diffraction grating in 2005. This shift was corrected on the basis that the ChlF for all weed species or corn hybrids is due to chlorophyll-*a* fluorescence only. Therefore, the mean location of the far-red fluorescence (FRF) peak was forced to 735 nm by applying a uniform wavelength shift for each year. After shifting, the spectra were truncated to the 400 nm to 755 nm range.

All data processing has been performed using custom scripts and the PLS_Toolbox¹¹ version 5.2.2 under MatlabTM and custom scripts and packages under the statistical computing

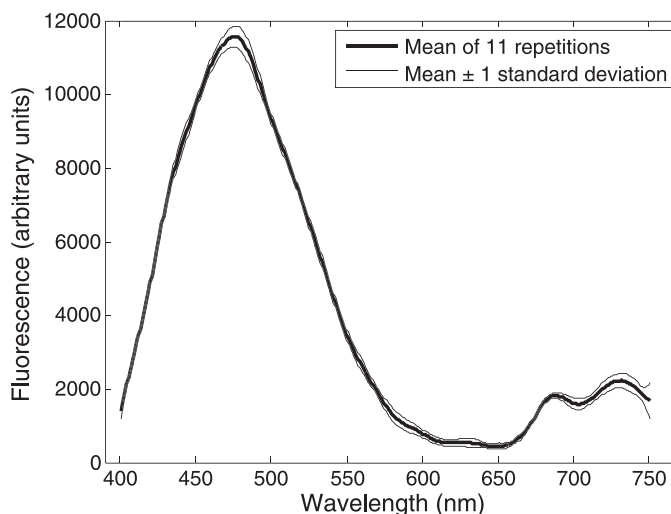


FIG. 2. Mean and standard deviation around the mean for 11 raw spectra. *Panicum capillare* (L.) at the 15–20 days stage, apex of leaf blade.

environment R.¹³ Unless otherwise stated, default options were retained when using the PLS_Toolbox and R packages.

RESULTS

Mean Fluorescence Emission Spectra. In a first step, fluorescence spectra were averaged to observe the effects of growth stage, position on the leaf blade and year. For both corn and monocots, as the plants were aging from 15–20 days to 25–30 days, the intensities of the BGF increased by ~30% to ~40%, whereas the intensities of ChlF decreased to a lower extent (Fig. 3). Measurements made at different leaf positions also influenced the fluorescence emission spectra, although the effect was less marked than that for plant age (Fig. 4). In general, both BGF and ChlF intensities measured at the leaf apex were lower by about 10% than those measured at the leaf base.

As mentioned above, variability between the different data sets (2005, 2006, and 2007) was introduced by cultivating plants and by measuring fluorescence emission spectra under

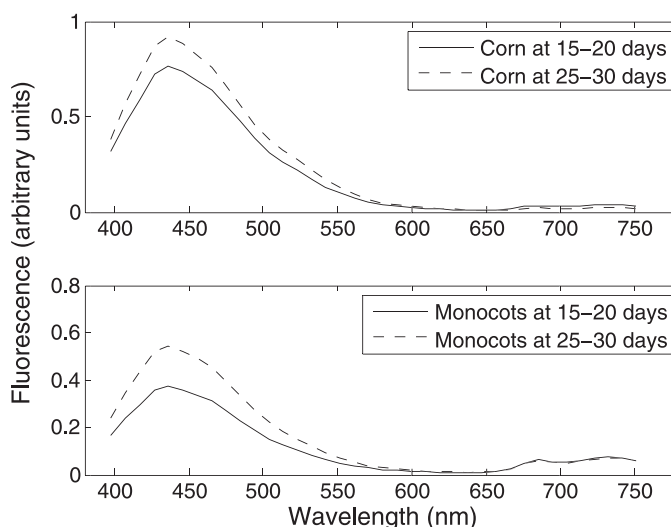


FIG. 3. Raw fluorescence spectra averaged by plant group and growth stage.

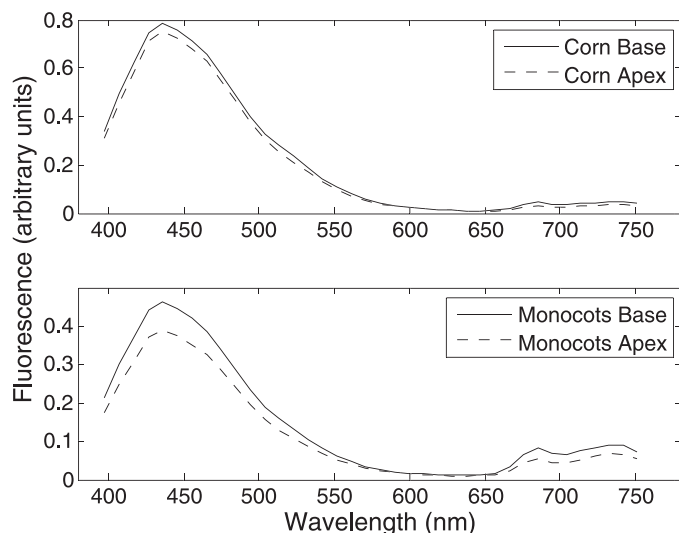


FIG. 4. Raw fluorescence spectra averaged by plant group and position on the leaf blade.

different environmental conditions. Despite the large variations observed between the different data sets, it can be seen that in general, corn can be distinguished from monocot weeds by its higher BGF intensities and lower ChlF intensities (Fig. 5). However, there are important differences between the three data sets. The BGF intensity of corn in 2007 was similar to the BGF intensity of monocot weeds in 2006. Also, the ChlF intensity of monocot weeds in 2006 was similar to that of corn measured in 2005 and 2006, whereas the ChlF intensity of corn in 2007 was similar to the ChlF intensities of monocot weeds in 2005 and 2007.

From observation of the mean spectra, it can be concluded that the effect of year on fluorescence variability dominated, followed by the effects of leaf age and leaf position. The year-to-year variation within a plant group was of the same order of magnitude as the variation between plant groups within a single year.

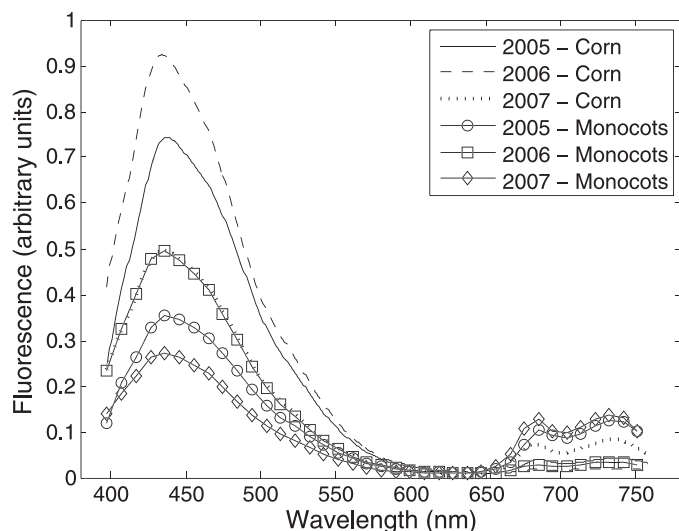


FIG. 5. Raw fluorescence spectra averaged by plant group and year. Note that in the BGF range, the mean spectra for 2007-Corn overlaps that for 2006-Monocots.

TABLE III. PLSDA models on the whole spectra for corn/weed monocot discrimination. Row headings identify the calibration data set. Numerical column headings identify the test data set. The errors in cross-validation (in bold) and prediction are given in percentage.

	No. of latent variables	2005	2006	2007
2005	4	4.8	7.3	42.2
2006	4	7.7	3.5	40.1
2007	3	32.9	46.2	9.6

Discrimination Based on the Full Spectrum. To discriminate monocot weeds from corn, a PLSDA^{11,14} model based on the complete spectra was developed for each of the three years. Based on previous work,¹⁰ the spectra were divided by the sum in the 570–620 nm band ($F_{570-620}$) for normalization. Cross-validation was performed using the Venetian Blind¹¹ method with 10 splits. In the PLSDA models, the number of latent variables was automatically selected by the software. The yearly PLSDA models were applied to the data sets from the other two years and the prediction errors were computed (Table III). In all cases, the cross-validation error was smaller than 10% and the prediction errors were large when the 2007 data set was involved either as a calibration or as a prediction data set. For these cases, the mean prediction error was 40.4%. This is close to a 50% error rate that corresponds to pure chance for a two-class model. The regression coefficients (Fig. 6) for the 2005 and 2006 models were similar but the ones for 2007 were different. The loading vectors were similar for the three years up to the second latent variable only (data not shown). Models limited to two latent variables were calibrated for each year. Based on the observation of the loading vectors, it was expected that models limited to two latent variables would generalize better. For these models, the cross-validation errors increased to about 15% but the prediction errors when the 2007 data was involved (calibration and prediction) remained high, in the 30% to 46% range.

Model Transfer. The prediction error could be reduced by applying model transfer schemes. Model transfer amounts to adapting the model calibrated on a primary data set to make it

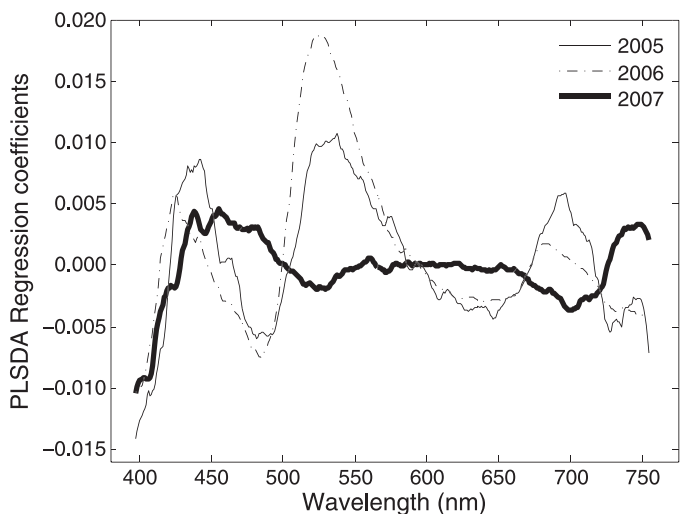


FIG. 6. Regression coefficients from the PLSDA model based on full spectra normalized by $F_{570-620}$. Two-class model: monocot and corn.

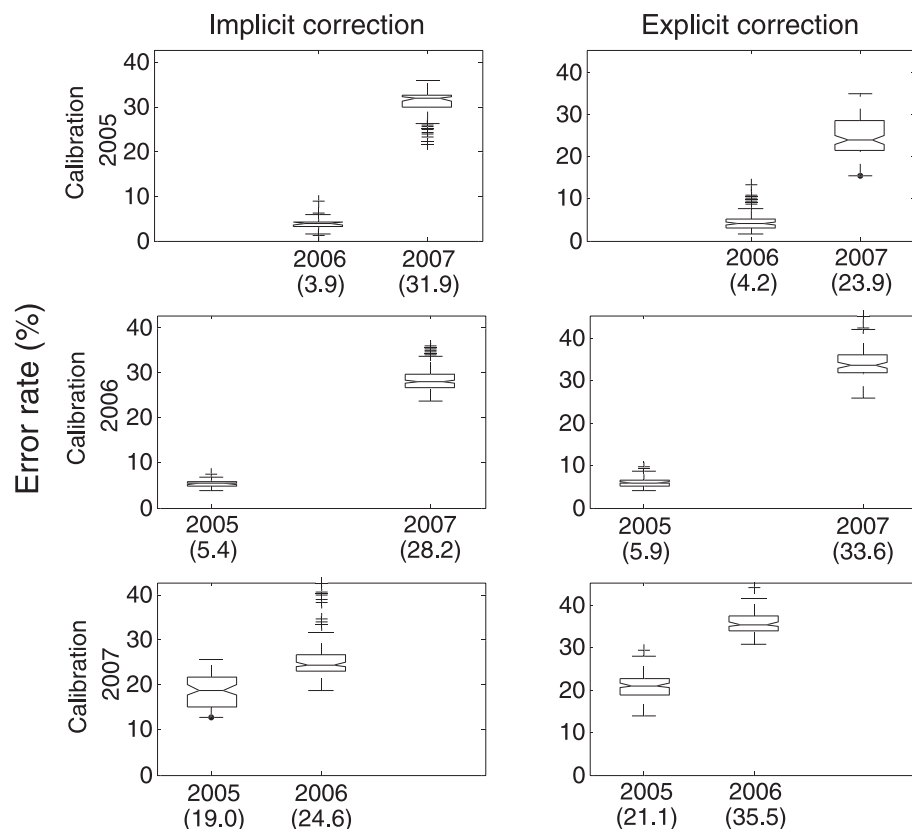


FIG. 7. Box and whisker plots of the prediction error for implicit and explicit model transfer approaches. Numbers in parenthesis are the median value for the corresponding box.

work on another data set (secondary). Ideally, the differences between the primary and the secondary data sets are small and weakly correlated to the discriminating factors embedded in the data. Model transfer methods can be grouped under two main approaches:¹⁵

- (1) Implicit model correction: based on the reconstruction of the model using a calibration data set made of spectra from both the primary and secondary sets.
- (2) Explicit model correction: first, the differences between the primary and secondary data sets are identified. Then, the model is recomputed from the primary data set modified so it “resembles” the secondary data set. This modification can be achieved directly on the model using orthogonal projection of the difference.^{16,17} Alternatively, the model is not modified but spectra from the secondary data set are preprocessed to make them compatible with the ones from the primary set (e.g., optical standardization¹⁸).

Both approaches were used to transfer PLSDA models from one year to the other two years for a total of six transfers. To perform the transfers, 30 corn and 30 monocot spectra from the secondary year were used. These 60 samples were selected at random. The use of 30 samples represented at most 29.4% of the population (monocots in 2007). For the implicit correction, the 60 spectra were added to the primary data set and the PLSDA model was recomputed and tested on the secondary year. For the explicit correction, 30 corn and 30 monocot spectra from the primary year were picked at random and subtracted from the corresponding spectra (i.e., corn from corn and monocot from monocot) in the secondary year. A principal

component analysis (PCA) of this difference matrix was performed. The k first components were subtracted from the primary data set by orthogonal projection of this set onto the space defined by the k components.¹⁶ The new PLSDA model was built from this reduced primary data set and tested on the secondary data set. With this procedure, the prediction error was a function of both k and the number of latent variables. The final model corresponds to the combination of these two factors giving the smallest prediction error. When there was no clear minimum (monotonically decreasing error), the Cattell Scree rule¹⁹ was applied, first on the number of latent variables at a fixed k and then on k . As the model transfer required random sampling, the process was repeated 50 times, yielding distributions of prediction errors, number of latent variables, and k values.

Most of the time, the PLSDA models had about five to six latent variables and optimum k values were 3 or 4 (data not shown). Both model transfer approaches yielded similar results. In all cases, the model transfer reduced the prediction errors (Fig. 7). However the prediction error for cases involving the 2007 data set either as a calibration or as a prediction data set remained high with median values ranging between 19% and 36%.

Identification of Discriminant Regions in the Spectra. In order to improve the robustness of the discrimination models (i.e., decrease the prediction error), we further examined the structure of the fluorescence spectra by calculating an autocorrelation matrix from all available raw spectra. The coefficients are plotted as an image in Fig. 8. The intensity of a point corresponds to the correlation coefficient of the

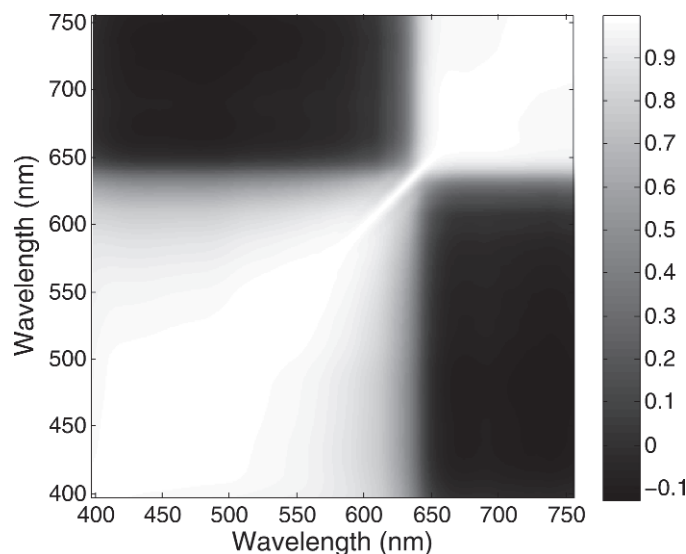


FIG. 8. Autocorrelation matrix from raw spectra.

fluorescence intensity at two wavelengths. There are two well-defined wavelength ranges of high correlation: $F_{400-600}$ or the blue-green fluorescence range (BGF) and $F_{650-755}$ or the chlorophyll fluorescence range (ChlF). The correlation between these two ranges is small. These are connected by a transition range ($F_{600-650}$). In this transition range, part or all of the correlation can be attributed to optical diffusion (diffraction and dispersion) and numerical diffusion resulting from the use of smoothing filters.

Because the two domains composing the fluorescence spectra may vary independently, they may contribute to different extents to the resulting prediction error (robustness) of the PLSDA models. Recently, robust discrimination between monocots and dicots (two classes) has been achieved by using BGF alone.⁴ Therefore, the potential of BGF to discriminate corn hybrids and monocot weeds was tested by calculating PLSDA models (Table IV). The algorithm had to select up to five latent variables to achieve the classification as determined using cross-validation within a single year. This number of latent variables was higher than the one expected from the autocorrelation matrix (Fig. 8). The models calibrated from yearly data performed poorly in prediction with results close to a random assignment (50% error rate). Clearly, PLSDA cannot create two well-defined classes.

This result was obtained because in the BGF region of the spectrum, the monocots, including corn hybrids, form a homogeneous group. This is clearly shown when the BGF spectra are normalized by the mean intensity in the band $F_{415-455}$ centered on the BGF peak at F_{435} (Fig. 9). On a yearly

TABLE IV. PLSDA models on the 400–550 nm band for corn/weed monocot discrimination. Row headings identify the calibration data set. Numerical column headings identify the test data set. The errors in cross-validation (in bold) and prediction are given in percentage.

	No. of latent variables	2005	2006	2007
2005	4	23.4	42.8	51.0
2006	5	40.2	20.4	43.9
2007	3	49.0	45.0	23.6

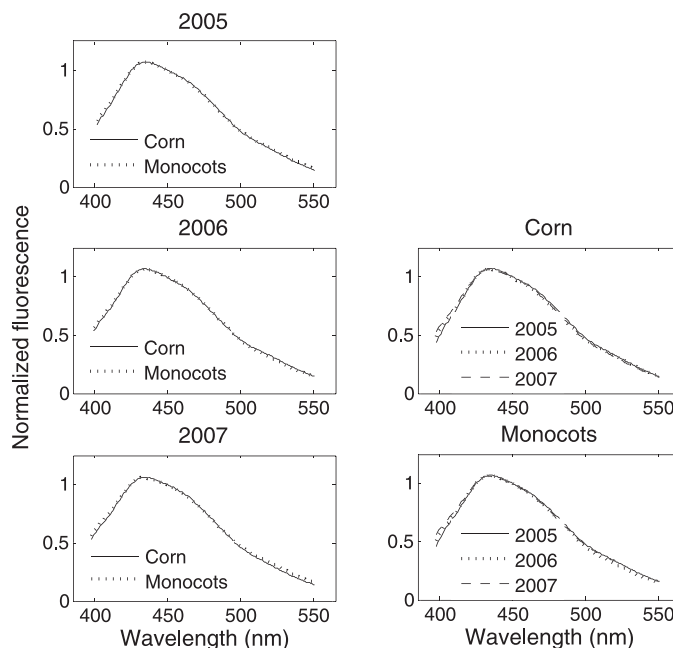


FIG. 9. Mean BGF spectra for the two groups and the three years, normalized by $F_{415-455}$.

basis, the mean BGF spectra of corn hybrids cannot be distinguished from the mean spectra of monocots. There were more differences between years within a single class (right column of Fig. 9) than there were between classes in a single year. From these results, it was concluded that the BGF spectrum shape does not include any useful information for discriminating between corn hybrids and monocots. Therefore, in accordance with our previous findings,¹⁰ the BGF was used for normalization leaving only normalized ChlF as a discriminating signal.

Discrimination Based on Chlorophyll Fluorescence Spectra. Based on the above conclusion, the mean chlorophyll fluorescence spectra were normalized by the $F_{415-455}$ intensity for corn hybrids and monocot weeds. The normalized ChlF spectra were different on a yearly basis (Fig. 10). At all wavelengths, ChlF was lower for corn than for monocot weeds. However, the year-to-year variation in the normalized ChlF for corn or monocots was of the same order of magnitude as the difference between corn and monocots on a yearly basis. The results from PLSDA modeling confirmed these observations (Table V). The cross-validation error was approximately constant at 15%. This was higher than the cross-validation error for models based on the whole spectra (Table III). Compared to the cross-validation errors, the prediction errors were much higher, indicating that the PLSDA models based on the normalized ChlF lacked robustness.

It is noteworthy that in these models, the number of latent variables was low. Moreover, the first two latent variables were similar for the three years. As an example, the ones from the model built from the 2005 data set are shown in Fig. 11. They are easy to interpret: the first latent variable is the sum (or integral) of the whole ChlF, while the second one computes the difference between F_{735} and F_{685} levels. By plotting for each year the corresponding values of the two latent variables (Fig. 12), it became clear that most of the discrimination was performed along the first variable (ChlF intensity). PLSDA

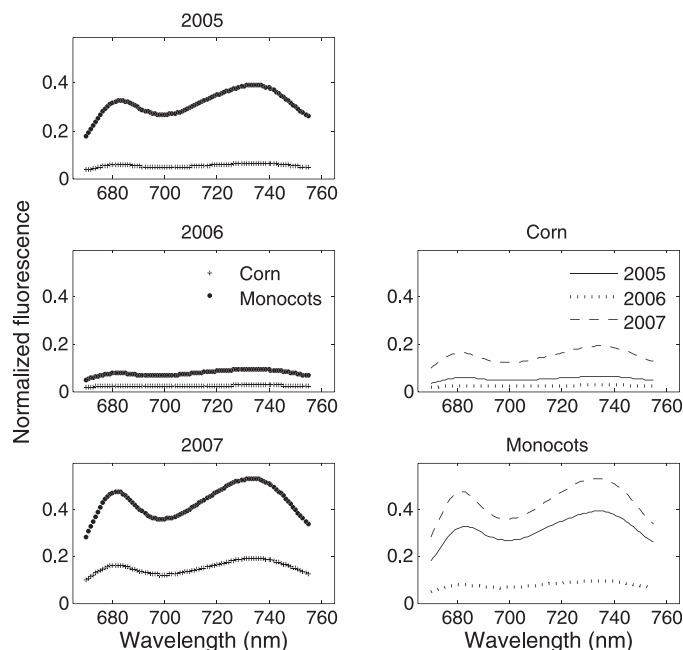


FIG. 10. Mean ChlF spectra for the two groups and the three years, normalized by $F_{415-455}$.

models including only this first latent variable gave higher cross-validation errors (Table VI) compared to models using the optimum number of latent variables (Table V), but the prediction errors were similar. As the data have previously been normalized with $F_{415-455}$, the model built from the sum over $F_{670-755}$ was equivalent to a ChlF/BGF ratio that was reported as a discriminating factor for plant groups.¹

Minimizing the Prediction Error Based on Normalized ChlF Intensity. The results presented above indicated that the mean ChlF intensity ($F_{670-755}$) normalized by the BGF intensity ($F_{415-455}$) contains most of the discriminant information, but PLSDA models based on this variable lack robustness. An alternative approach is therefore needed. To this end, using a Gaussian kernel¹³ we computed the probability densities for the mean ChlF ($F_{670-755}$) after normalization by $F_{415-455}$ for each plant class and each year (Fig. 13). The densities have a similar shape but the scales are different. On a single dimension, the discrimination between two classes (here corn and monocots) reduces to setting a threshold. In our case, samples with a normalized ChlF value below the threshold were assigned the corn label; others were classified as monocots. The classification error rate depended on the threshold value; it had a well-defined minimum (Fig. 14). It was approximately constant at 15% from year to year (Table VII). Looking at the false positive and false negative rates (data

TABLE V. PLSDA models on the 670–750 nm band for corn/weed monocot discrimination. Row headings identify the calibration data set. Numerical column headings identify the test data set. The errors in cross-validation (in bold) and prediction are given in percentage.

	No. of latent variables	2005	2006	2007
2005	2	15.8	41.0	29.9
2006	2	23.3	15.4	44.7
2007	1	30.1	48.9	15.7

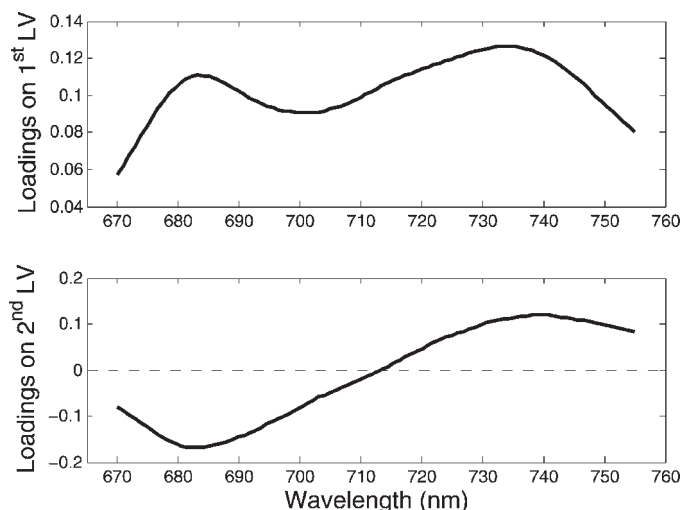


FIG. 11. Loadings of the PLSDA model on $F_{670-750}$ for corn/weed monocot discrimination built from the 2005 data set normalized by $F_{415-455}$. First and second latent variables (LV).

not shown) confirmed that the classification error was not biased towards corn or monocots. This result is similar to the performance of the PLSDA on the same spectrum region (Table V). However, the associated thresholds varied from year to year.

A method to set the threshold automatically for each year is necessary. This can be achieved based on the cumulative probability density of the corn population. Corn was selected because in any given corn field, it is much more straightforward to sample corn plants than to sample the monocot weed population, which may be diversified and unevenly distributed across the field. The overall process is illustrated for the 2005 data set (Fig. 15). Starting from the minimum error rate (a), the optimal threshold value (b) is identified as well as the corresponding fraction on the cumulative distribution of ChlF intensity (c). These values were comparable for the three years (Table VII) with an average of 85%. Therefore, a threshold value (d) corresponding to 85% quantile (Q_{85}) could be used. These thresholds were slightly different from the optimal ones, but the error rate (e) remained close to the minimum (a) (Table VII). Using the above technique to set up a threshold resulted in a discrimination process that is robust, with an error rate varying from 14.6% to 16.8%.

Setting up the threshold requires that the normalized ChlF intensity (i.e., the ChlF/BGF ratio) be measured on a sample of corn plants. The sample should be large enough to yield a reasonably accurate estimate of the 85% quantile. The required sample size was estimated using bootstrapping. For each yearly data set, samples of varying size made of normalized ChlF intensity were picked at random. For each sample, the 85% quantile point was estimated using the function “quantile” in R.¹³ For a given sample size, the process was repeated 1000 times yielding a vector of 1000 estimates of Q_{85} . Bootstrapping was applied to this vector to estimate the median and the 90% confidence interval. The function “boot” in R¹³ was used.^{20,21} The confidence intervals were compared to threshold bounds to define the acceptable sample size. Threshold bounds were defined as the lower and higher limits of threshold resulting in an error rate within 10% of the error rate at Q_{85} . The results (Fig. 16) showed that the 90% confidence interval was completely included within the threshold bounds for sample

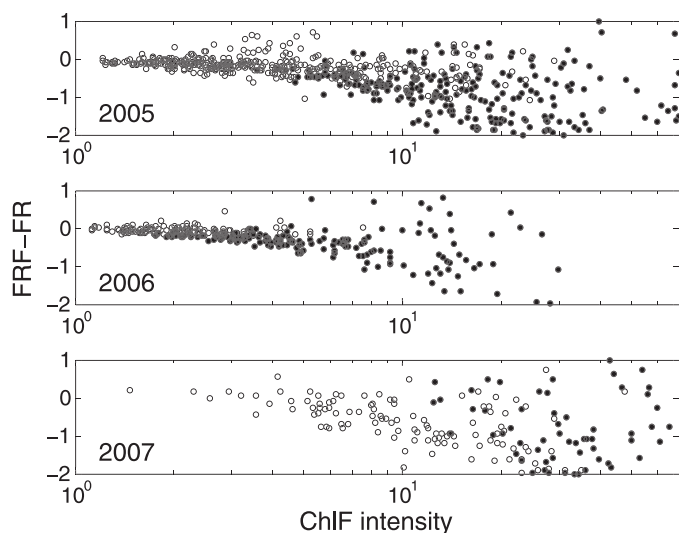


FIG. 12. For each year, ChlF intensity (i.e., the sum of the ChlF ($F_{670-755}$)) and the difference between FRF ($F_{725-745}$) and RF ($F_{675-695}$) for the corn (○) and monocots (●).

sizes of about 65, 45, and 75 for years 2005, 2006, and 2007, respectively. Therefore, a sample size of 75 is recommended.

DISCUSSION AND CONCLUSION

The objectives of this study were to identify the most relevant information contained in the plant UV-induced fluorescence emission spectra from which a model could be developed to discriminate monocot weeds from corn hybrids. This model needs to be accurate and robust but still easy to build to allow its implementation in the field.

The results demonstrated that PLSDA models based on the

TABLE VI. PLSDA models on $F_{670-750}$ for corn/weed monocot discrimination using only one latent variable. Row headings identify the calibration data set. Numerical column headings identify the test data set. The errors in cross-validation (in bold) and prediction are given in percentage.

	% variance captured	2005	2006	2007
2005	99.2%	19.2	42.9	22.5
2006	98.5%	20.7	21.8	46.3
2007	98.4%	30.1	48.9	15.7

entire spectra of UV-induced fluorescence could efficiently discriminate corn from monocot weeds, with cross-validation error varying between 3.5% and 9.6%. This indicates that the PLSDA models based on full spectra of UV-induced fluorescence can efficiently cope with intra-set variations introduced by measurements made on different leaf positions and different leaf ages, in agreement with previous results⁸ obtained with other modeling approaches.

The use of plants having two different ages but still relatively young (less than 30 days old) was justified by the application of herbicides in corn fields, which usually occurs during the first month after emergence. The observed increase of BGF intensities and decrease of ChlF observed in older plants (25–30 days old) relative to younger plants (15–20 days old) are coherent with published data,⁹ where there was a 36% increase and a 37% decrease of BGF and ChlF, respectively, in wheat leaves aging from 10 to 15 days old. In that study, the BGF increase during leaf aging was shown to result from the accumulation of large sclerenchyma bands enriched in ferulic acid, the main emitter of BGF.⁵ Concerning the ChlF, its decline during leaf aging was due to the accumulation of non-fluorescent UV-absorbing metabolites in leaf epidermis, which decrease the UV excitation of chlorophyll molecules located in leaf mesophyll.

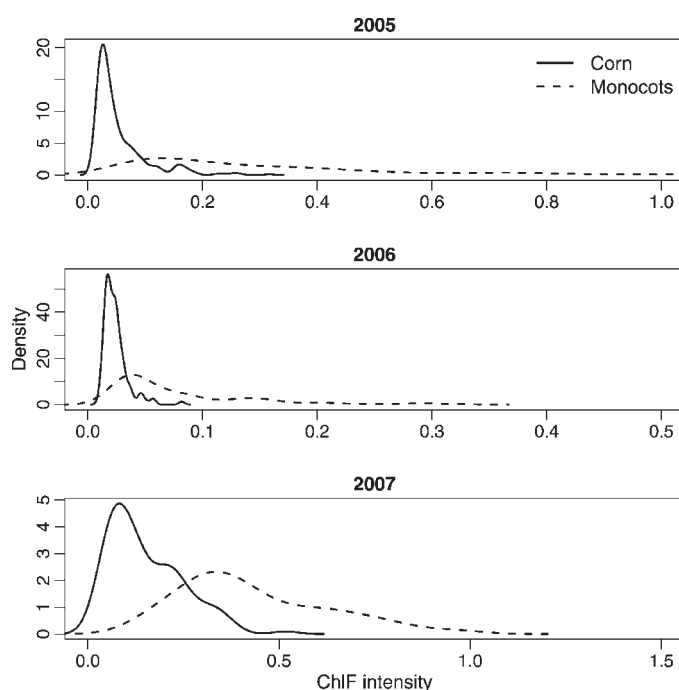


FIG. 13. Estimated probability densities of the ChlF intensities normalized by BGF intensities for each plant group and each year.

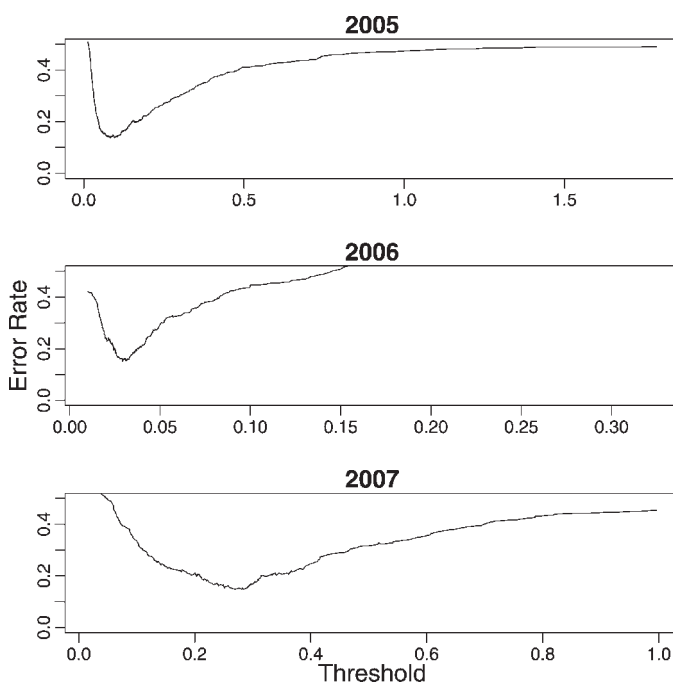


FIG. 14. Classification error rate versus threshold of normalized ChlF intensity for corn/monocot discrimination. Note that the horizontal scales differ from plot to plot.

TABLE VII. For each year, the information extracted from the ChlF intensity probability densities of the corn population to build the classifier for corn/weed monocot discrimination. [Bold letters refer to Fig. 15.]

	2005	2006	2007
Minimum error (%), [a]	14.2	15.8	15.2
Optimal threshold, [b]	0.082	0.031	0.289
Quantile at optimal threshold, [c]	0.83	0.84	0.89
Threshold at Q_{85} , [d]	0.089	0.032	0.263
Error (%) at Q_{85} , [e]	14.6	16.8	15.8

Whereas the cross-validation errors of PLSDA models calibrated on yearly data sets were less than 10%, the prediction errors computed when applying yearly models to classify data from another year could be as high as 46%. The prediction error was better between years 2005 and 2006 but degraded when the 2007 data set was involved either as the calibration or as the test set. For both corn and monocot weeds, BGF intensities and ChlF intensities were the highest and the lowest respectively in 2006, whereas the opposite was observed in 2007. These differences could be related to the environmental conditions during plant growth and/or fluorescence measurements, which took place in a greenhouse in 2007 and in a growth chamber in 2006. In 2005, plants were grown in a growth chamber (as in 2006) and fluorescence was measured in a greenhouse. Therefore, the growth conditions are responsible for much of the variations in the fluorescence spectra that masked the discriminant information and increased the prediction errors. At this point, without further information it is difficult to explain the differences between the spectra of 2006 and 2007. By their lowest BGF and highest ChlF intensities, it seems that the leaves of plants grown in the greenhouse in 2007 contained less ferulic acid in the cell walls and less UV-absorbing metabolites (e.g., soluble flavonoids) in the vacuoles of epidermal cells. Lower concentrations of these secondary metabolites may be explained by different factors such as, for example, lower growth light intensities, slower leaf development resulting in physiologically younger leaves during fluorescence measurements,⁹ and non-limiting nitrogen supply.^{22,23}

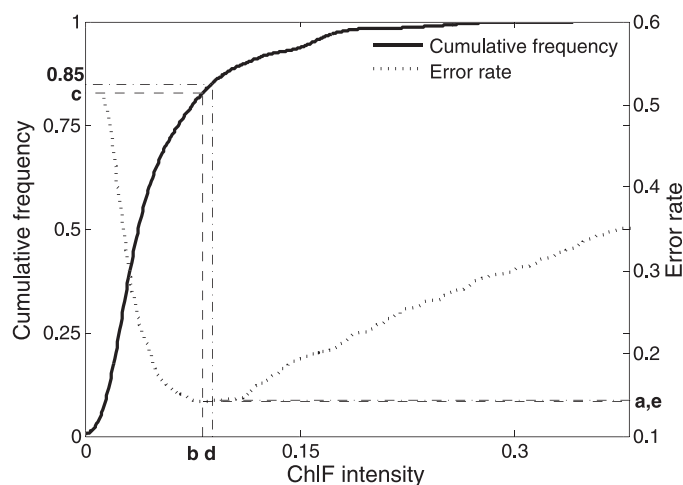


FIG. 15. Threshold selection and classification error associated with a fixed threshold based on Q_{85} for the cumulative frequency curve for corn. 2005 data set. **a**: Minimal error rate; **b**: optimal threshold; **c**: quantile at **b**; **d**: threshold at Q_{85} ; and **e**: error rate at **d**.

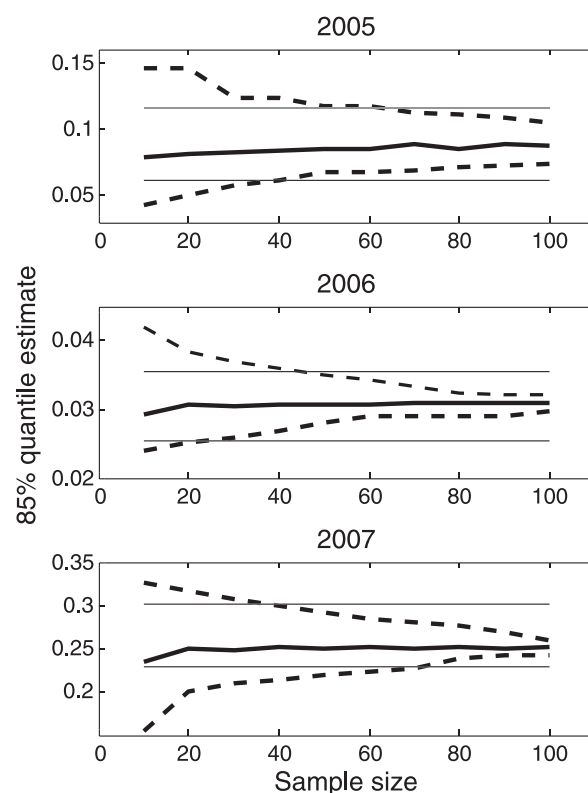


FIG. 16. Effect of sample size on the accuracy of the threshold estimate. ((Bold solid line) Median of Q_{85} estimates; (bold dashed lines) 90% confidence interval on Q_{85} ; (thin solid lines) threshold bounds). Intersection between 90% confidence interval curves and threshold bounds gives required sample size.

The results lead to the conclusion that for discriminating monocot weeds from corn, the ratio ChlF/BGF is the best factor available. A similar discriminating factor was reported in the literature for discriminating between monocots and dicots.¹ In all our measurements (here and elsewhere^{8,10,24}), corn often displayed a smaller ChlF/BGF ratio than all monocot or dicot weeds. It is likely that corn leaves contain more ferulic acid in the cell walls and more UV-absorbing metabolites (e.g., soluble flavonoids) in the vacuoles of epidermal cells than leaves of monocot plants.

It has been shown that the ratio ChlF/BGF was the best discriminating variable but that the threshold to discriminate corn from monocots varied from data set to data set. The method to allow for straightforward recalibration of a discrimination model (i.e., setting the threshold) requires that the ChlF/BGF ratio be measured on a sample of corn plants and that the 85% quantile of the sample be estimated and used as a threshold between corn and monocot weeds. The required sample size for this estimation was found to be 75. In practice, recalibration should be performed each time growth conditions or corn hybrids change significantly. Measurement of the ChlF/BGF ratio can be performed with a fairly simple handheld instrument using a modulated UV diode for excitation, two wide-band detectors ($F415-455$ and $F670-755$), and phase-lock loop circuitry to isolate the fluorescence signal. With such an instrument, measurements on 75 corn plants can be performed in a few minutes. The main issue that cannot be resolved from the available data sets is the sampling pattern (field area covered and distribution of sampled corn plants in this area) that should be implemented.

For a sample size of 75 plants, the expected error rate should be better than 18.5%. This is the worst error rate at Q_{85} (16.8% in Table I) multiplied by 1.1 to suit the 10% increase associated with the threshold bounds (Fig. 16 and associated text). As the error is not biased towards corn or monocot weeds, there is equal chance of misclassification for corn and monocot. Classifying corn as weeds results in application of herbicides where they are not required. This is a loss in herbicides. On the other hand, classifying weeds as corn prevents triggering herbicide release where it would be necessary. While this can be acceptable for preserving crop yield, it may result in a weed buildup for the following years as a few untreated weeds can release a large amount of seeds.²⁵ This should be evaluated carefully before implementing a weed-detection system for site-specific weed management in corn.

ACKNOWLEDGMENTS

The excellent technical support of Mr. G. St-Laurent and Mrs. M. Piché is hereby acknowledged. This work was financially supported by Agriculture and Agri-Food Canada and by a Natural Sciences and Engineering Research Council of Canada Discovery Grant.

1. E. W. Chappelle, F. M. Wood, W. W. Newcomb, and J. E. McMurtrey III, *Appl. Opt.* **24**, 74 (1985).
2. C. Buschmann and H. K. Lichtenthaler, *J. Plant Physiol.* **152**, 297 (1998).
3. Z. G. Cerovic, G. Samson, F. Morales, N. Tremblay, and I. Moya, *Agronomie* **19**, 543 (1999).
4. B. Panneton, S. Guillaume, J. M. Roger, and G. Samson, *Appl. Spectrosc.* **64**, 30 (2010).
5. H. K. Lichtenthaler and J. Schweiger, *J. Plant Physiol.* **152**, 272 (1998).
6. P. J. Hilton, "Laser-induced fluorescence for discrimination of crops and weeds", in *Proceedings of the SPIE* (2000), p. 223.
7. A. A. Gitelson, C. Buschmann, and H. K. Lichtenthaler, *Remote Sens. Environ.* **69**, 296 (1999).
8. L. Longchamps, B. Panneton, G. Samson, G. D. Leroux, and R. Thériault, *Precision Agric.* **11**, 181 (2010).
9. S. Meyer, A. Cartelat, I. Moya, and Z. G. Cerovic, *J. Exp. Botany* **54**, 757 (2003).
10. B. Panneton, J. M. Roger, S. Guillaume, and L. Longchamps, *Appl. Spectrosc.* **62**, 747 (2008).
11. B. M. Wise, N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig, and R. S. Koch, *PLS_Toolbox 4.0 - Reference Manual for use with Matlab* (Eigenvector Research Inc., 2006).
12. J. H. Norikane and K. Kuruta, *Trans. ASAE* **44**, 1915 (2001).
13. R. Development Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2009).
14. M. Barker and W. Rayens, *J. Chemom.* **17**, 166 (2003).
15. P. Gujral, M. Amrhein, and D. Bonvin, *Anal. Chim. Acta* **642**, 27 (2009).
16. J. M. Roger, F. Chauchard, and V. Bellon-Maurel, *Chemom. Intell. Lab. Syst.* **66**, 191 (2003).
17. A. Andrew and T. Fearn, *Chemom. Intell. Lab. Syst.* **72**, 51 (2004).
18. Y. Wang, D. J. Veltkamp, and R. Kowalski, *Anal. Chem.* **63**, 2750 (1991).
19. G. Saporta, *Probabilités, Analyse des données et statistiques* (Editions Technip, 1990).
20. A. Canty and B. Ripley, *boot: Bootstrap R (S-Plus) Functions*. R package version 1.2-36 (2009).
21. A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Applications* (Cambridge University Press, Cambridge, 1997).
22. A. Cartelat, Z. G. Cerovic, Y. Goulas, S. Meyer, C. Lelarge, J.-L. Prioul, A. Barbotin, M.-H. Jeuffroy, P. Gate, G. Agati, and I. Moya, *Field Crops Res.* **91**, 35 (2005).
23. S. A. Mercure, B. Daoust, and G. Samson, *Can. J. Botany* **82**, 815 (2004).
24. B. Panneton, L. Longchamps, R. Thériault, and G. D. Leroux, "Fluorescence spectroscopy of vegetation for weed-crop discrimination" (ASABE Annual International Meeting, Portland, OR, July 9–12, 2006), Paper No. 061063.
25. M. J. Simard, B. Panneton, L. Longchamps, C. Lemieux, A. Légère, and G. D. Leroux, *Weed Science* **57**, 187 (2009).