Multivariate analysis of laryngeal fluorescence spectra recorded in vivo

Eker, Charlotta; Rydell, Roland; Svanberg, Katarina; Andersson-Engels, Stefan

Published in:
Lasers in Surgery and Medicine

DOI:
10.1002/lsm.1048

2001

Citation for published version (APA):
Multivariate Analysis of Laryngeal Fluorescence Spectra Recorded In Vivo

Charlotta Eker, PhD,1 Roland Rydell, MD, PhD,2 Katarina Svanberg, MD, PhD,3 and Stefan Andersson-Engels, PhD1*

1Department of Physics, Lund Institute of Technology, SE-221 00 Lund, Sweden
2Department of Oto-Rhino-Laryngology, Lund University Hospital, SE-221 85 Lund, Sweden
3Department of Oncology, Lund University Hospital, SE-221 85 Lund, Sweden

Background and Objective: The potential of using various multivariate analysis methods for classification of fluorescence spectra acquired in vivo from laryngeal tissues in Patients was investigated.

Study Design/Materials and Methods: Autofluorescence spectra were measured on 29 normal tissue sites and 25 laryngeal lesions using 337-nm excitation. Four different multivariate analysis schemes were applied. Laryngeal fluorescence spectra from patients who had been administered δ-aminolevulinic acid (ALA) were obtained using 405-nm excitation and were classified using partial least squares discriminant analysis (PLS-DA).

Results: For autofluorescence spectra, logistic regression based on principal component analysis (PCA) or PLS, or PLS-DA all resulted in sensitivities and specificities around 90% for lesion vs. normal. Using ALA and 405-nm excitation gave a sensitivity of 100% and a specificity of 69%.


Key words: in vivo diagnosis; larynx; laser-induced fluorescence; partial least squares (PLS); spectroscopy; tissue characterization; multivariate analysis

INTRODUCTION

The aims of this study were twofold: one was to investigate whether fluorescence can be used to distinguish between normal laryngeal tissue and malignant or premalignant lesions to guide the biopsy sampling or demarcate lesion borders during surgery. The other was to evaluate which multivariate analysis tools can be used to classify the tissue from the recorded data. Various lesions on the vocal and ventricular folds (true and false vocal cords) have been examined by fluorescence emission spectroscopy. Both the tissue autofluorescence and the fluorescence from δ-aminolevulinic acid (ALA)-induced protoporphyrin IX (PpIX) as a fluorescent tumor marker have been investigated. Different multivariate methods have been explored, and the results are compared. The characterization of tissues and diagnosis of diseases are very much based on histopathologic methods. These are fairly reliable and comparative studies show a good agreement between the judgement of different pathologists evaluating the same biopsy specimens [1,2]. However, a disadvantage is that it can only be applied in vitro, i.e., a small tissue sample (biopsy) has to be removed. The results are usually not available until a few days after the examination. Novel methods to be used complementary to conventional biopsies, helping the physician characterize lesions in real time during an examination, are, thus, under development and evaluation. Such methods could be used to help distinguish the lesion borders during surgical removal or to help the endoscopist find the most relevant sites for biopsy sampling. Both of these aspects are very important in handling laryngeal lesions, because this organ is small and delicate and removal of even small tissue samples could jeopardize its function.

The possibility of using laser-induced fluorescence (LIF) for classification of different types of biologic tissue has been studied extensively for many years [3–5]. The main objectives are that LIF measurements are safe and noninvasive and can be performed in real time during a clinical examination. Several studies have aimed at evaluating the potential of various fluorescence techniques to identify premalignant or malignant lesions of the larynx [6–13]. The results from the reported studies vary, and it still has to be determined how to best characterize laryngeal lesions using fluorescence methods and to evaluate whether such a method could be of clinical interest. A recent review paper describes the published fluorescence studies in the oral cavity [14].

Some evaluation criteria must be chosen to classify the fluorescence spectra, e.g., as originating from diseased or healthy tissue. This is a challenging task, because the biologic variation is very large and the fluorescence spectra complex due to the many different fluorophores.
contributing as well as influence from absorption and scattering of the light inside the tissue. There are two classes of methods available for this purpose: the modeling approach and the statistical approach. The modeling approach aims at quantifying the contribution to the fluorescence spectra from different tissue chromophores. The insight in this relationship between chromophore concentrations and spectral information can be used to discriminate between diseased and normal tissues. The second type of method, the statistical approach, is more empirical but often clinically effective and easy to implement. Raw fluorescence data are in this case used to establish a model to distinguish between diseased and normal tissue. In this study, we explore the potential of a few statistical multivariate methods to classify fluorescence spectra from tissue.

In tissue spectroscopy, multivariate linear regression (MVLR) is the most extensively used of the multivariate methods [15–18], but principal component analysis (PCA) and partial least-squares (PLS) analysis have also been used [19–21]. MVLR may fail if there are more variables than samples and if there are collinearities in the data. To avoid these problems, some variables in the spectral data have to be deleted. Stepwise regression is a method of choosing some variables that do not covary. This approach works well in some but not all cases. However, when using PCA- or PLS-based methods, data of full rank can be used. These methods are at least as powerful and simple to use as MVLR. Thus, we see no need to use MVLR for this application. However, it is not clear to us whether a PCA- or PLS-based technique is favorable for this type of classification of one homogeneous (normal tissue) and one heterogeneous group (premalignant or malignant lesions). Although PCA finds the components that account for most of the spectral variation, PLS is best suited to extract information that correlates with the concentration of a certain tissue constituent. Thus, it is not obvious that a classification into two discrete types, one of them being very heterogeneous, is optimally treated with PLS. Probability-based classification schemes such as logistic regression have been applied by some groups [22–24]. This is a nonlinear method that provides the probabilities of a sample belonging to the different diagnostic categories. This is basically a method that can be used as an adjunct to the other methods. It transforms the results to more easily interpreted data.

The results from an in vivo investigation of tissue autofluorescence and ALA-induced PpIX fluorescence for the characterization of lesions on the vocal and ventricular folds were evaluated. For the analysis, four different multivariate methods were used: soft independent modeling of class analogy (SIMCA) based on PCA, PCA classification combined with logistic regression, partial least squares discriminant analysis (PLS-DA), and PLS combined with logistic regression.

**MATERIALS AND METHODS**

LIF measurements on ventricular and vocal folds (false and true vocal cords) were performed in vivo with a fiber based optical multichannel analyzer (OMA) system [25] in connection with routine visual examination and biopsy sampling of suspect laryngeal lesions. The patients were under general anesthesia during the examination.

Forty-two patients (32 men and 10 women) were included in the study. The patients were between 23 and 87 years old, median age being 67.5 years. Twenty-one of the patients were given ALA by mouth as a fluorescent tumor marker in a dose of 7.5 or 5 mg/kg body weight approximately 4 hours before the examination. Two groups of data are presented: one from 405-nm excitation, including all patients who were given ALA, and one from 337-nm excitation, including the patients who were not given ALA and an additional five of the ALA patients, who were measured with both wavelengths. The PpIX absorption at 337 nm is very low; therefore, the administration of ALA will not affect the results at this excitation wavelength.

In total, 52 normal sites and 33 lesions from the vocal and ventricular folds are included in this analysis. After fluorescence measurement, all lesions were biopsied and classified by routine histopathology. The 33 lesions were classified as either moderate dysplasia, severe dysplasia, carcinoma in situ (CIS), or carcinoma. These groups of lesions are all malignant or premalignant and require removal.

In all, 26 normal sites, 3 dysplasias, 6 CIS, and 4 carcinomas were measured with 405-nm excitation. Twenty-nine normal sites, 8 dysplasias, 10 CIS, and 7 carcinomas were measured with 337-nm excitation. Of these, four normal, three dysplasias and three CIS were from ALA patients, and measured by both excitation wavelengths (except for one dysplastic lesion and one normal). The ambition was not to use fluorescence to separate the different lesion types, but rather to separate lesion from normal to assess lesion boundaries during endoscopy. Hence, all malignant and premalignant lesions were grouped together and simply called lesions in the following analysis. A few benign and low-grade dysplastic lesions were also measured, but not included in this study.

The light sources used for excitation were a nitrogen laser (Laser Science VSL-337 ND) operating at 337 nm and a nitrogen pumped dye laser (Laser Science DLM-220) using DPS as a dye, emitting at 405 nm. The excitation light was guided to the larynx through a 600-μm optical fiber, which was inserted through the endoscope and held in light contact with the tissue. The pulse energy out of the fiber was approximately 20 μJ at 337 nm and 1–2 μJ at 405 nm. The fluorescence light was collected by the same fiber and coupled out from the excitation path through a beam splitter. The fluorescence light was then focused onto the entrance slit of a polychromator (Acton SP-275). In front of the slit, cut-off filters (Schott GG385 and GG435 for 337- and 405-nm excitation, respectively) were placed to suppress scattered laser light. On the output, a time gated, image intensified CCD detector was placed, either a Peltier-cooled two-dimensional CCD array (Spectroscopy Instruments ICCD-576) or a linear diode array (Princeton Instruments ICCD-576).
Instruments TRY-512/L. The spectral resolution of the instrument was approximately 5 nm. The entire detection system was calibrated to a uniform spectral response using a tungsten lamp, and wavelength calibration was made using a mercury discharge lamp.

MULTIVARIATE METHODS

There are several different multivariate calibration methods available that can be used for spectral analysis \([26,27]\). PCA and PLS are two powerful chemometric methods, which both aim at decomposing the data into a structure and a noise part. The structure part, or model, is a linear combination of principal component spectra. Usually, it is sufficient to use a small number of principal components to characterize most of the spectral variations. Thus, the technique can be seen as a transform of the data into a new coordinate system with fewer dimensions. In PCA, the components are chosen for maximal explained variance, whereas PLS also takes into account the dependent variables (in this case, histopathology diagnoses). Hence, the PLS components describe only the variations in spectral shape that are most relevant for tissue pathology, whereas the impact of irrelevant spectral variations is reduced. To avoid confusion, in the following we will use the terms PLS components and principal components for the factors determined by PLS and PCA, respectively.

The data are represented in the same way for both methods. Each column in a matrix \(X\) represents one measured variable, in this case, the fluorescence intensity in a 3-nm-wide wavelength band for the various recorded sites, while each row in \(X\) corresponds to an observation, i.e., the fluorescence spectrum from 370- to 700-nm emission measured at one site. The multivariate calibration will result in a decomposition of \(X\):

\[
X = TP' + E,
\]

where \(P'\) is the transposed loading matrix, i.e., the matrix representing the principal component spectra, \(T\) is the score matrix containing the scores, i.e., the projections of all objects onto each principal component, and \(E\) is the residual matrix.

Before analysis, each spectrum was normalized by dividing all fluorescence intensities by the mean intensity of that spectrum and centered around zero intensity. The different multivariate analysis methods evaluated are briefly described below.

SIMCA classification. SIMCA classification was based on principal component analysis. A separate PCA model of each class (normal and lesion) was determined, and new samples were classified by looking at their distances from each class \([27]\).

PCA and logistic regression. PCA and logistic regression were used in combination for classification of fluorescence spectra. Spectral data were decomposed using PCA and reduced by only including the components that explain a certain amount of spectral variance. Logistic regression was then applied to calculate the posterior probability that a sample was either normal or a lesion. The model is based on Bayes theorem, stating that the probability of a sample belonging to the normal group (N), given a vector of principal component scores \((T)\) is

\[
P(N|T) = \frac{P(T|N) \times P(N) \times C(L|N)}{P(T|N) \times P(N) + P(T|L) \times P(L) \times C(N|L)},
\]

where \(P(N|T)\) is the posterior probability that the tissue is normal if its spectrum has the principal component scores \(T\), \(P(T|N)\) is the conditional probability that a normal tissue sample will have scores, \(T\), \(P(N)\) is the prior probability that the tissue is normal, and \(C(L|N)\) is the cost of misclassifying a normal tissue as lesion. \(C(L|N)\) and \(C(N|L)\) can be varied as long as their sum is equal to 1. Here, they were varied in steps of 0.05 to optimize the result, primarily for maximal number of correctly classified samples and secondarily for maximal sensitivity. The conditional probabilities were determined by fitting a normal probability density function to the scores for each principal component and each group (N and L), using the mean and standard deviation as free parameters in the fit. The conditional probability for all principal components were multiplied to yield the conditional joint probability for each group.

PLS and logistic regression. PCA looks for the principal components that account for most of the spectral variation. However, some of these components may be more or less irrelevant for the diagnosis of the tissue. Ramanujam et al. \([23]\) addressed this problem by determining the diagnostic content of each component using a Student’s \(t\)-test. The use of PLS analysis instead of PCA is another course of action. The principle of PLS is to find the components in \(X\) that describe as much as possible of the relevant variations in the spectra and at the same time have maximal correlation with the histopathology in \(y\), but without including the variations that are irrelevant or noise. All lesions were assigned a \(y\) value of \(-1\), and all normal tissue spots were assigned the value of \(+1\) when performing the PLS decomposition. The scores obtained were then used in the logistic algorithm to calculate the posterior probability of the tissue being normal, as described above.

PLS-DA The goal here was to use the information in \(X\) to predict the histopathologic properties of the tissue, which were represented by a column vector \(y\). When using PLS directly for discriminant analysis, the \(y\) variable was used as a binary code for the two groups. As before, all lesions were assigned a \(y\) value of \(-1\), and all normal tissue spots were assigned the value of \(+1\) when training the system. Based on cross-validation, a \(y\) value was predicted for each sample. A sample was then classified as normal if its predicted \(y\) was above a certain threshold. This threshold was moved in steps of 0.05 for optimal number of correctly classified samples. If more than one threshold gave the same total performance, the one giving maximal sensitivity was chosen.
Validation

Validation is always necessary to determine how many components to include in the model to avoid overfitting or underfitting and to study how the model will work for new data. In PCA, one looks at how much of the variance is explained by each additional principal component. Usually, only a few principal components will be sufficient to explain approximately 99% of the variance. Adding more components than that will not improve the model significantly, but possibly just will add noise to the model.

For PLS, validation produces a measure of the root mean square error of prediction (RMSEP), giving an estimate of how large an error we can expect when using the model to predict \( y \) of new data. The RMSEP is defined as the square root of the average of the squared differences between predicted and measured \( y \) values of the validation objects:

\[
RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}
\]

RMSEP is given in the original units of \( y \). Including more PLS components in the model will to a certain level give a better RMSEP. Beyond this level, adding more components will result in overfitting, i.e., including noise and irrelevant variations in the model. Hence, for both PCA and PLS, the number of components used was chosen by minimizing RMSEP.

For all algorithms described above except SIMCA, full cross-validation was used, where one sample at the time is left out and a model based on all other samples is used for prediction of the one left out. In a data set of \( n \) samples, \( n \) models based on \( n - 1 \) samples each are created for determining the prediction error. Full cross-validation is the best alternative when there are not enough samples to divide into separate calibration and test sets. For the SIMCA algorithm, the models were determined from separate calibration sets: one containing half of the normal and one containing half of the lesion spectra. Prediction was performed on a separate test set of data. All calculations were performed using a multivariate analysis software package (Unscrambler 7.5, Camo, Norway).

RESULTS

Autofluorescence From 337-nm Excitation

Normalized spectra from 29 normal (healthy) tissue spots, using 337-nm excitation can be seen to the left in Figure 1, and spectra from 25 malignant and premalignant lesions are to the right. The latter group consists of different types of lesions, from moderate and severe dysplasia to CIS and cancer. It can be seen in the figure that the fluorescence from healthy tissue has its highest peak around 390 nm, close to the fluorescence peak of collagen. However, in malignant or premalignant tissue, there is a more pronounced peak at around 460 nm, where NADH has its fluorescence peak. This is possibly due to a combination of effects: a thickening of the epithelial layer in the lesions, giving less pronounced collagen signal and a higher metabolic rate, which enhances NADH features, as well as hemoglobin signatures due to more vascularization. Thus, biochemical and morphologic features in the tissue can be detected by fluorescence measurements, and these can serve as indicators of cancerous or precancerous lesions.

SIMCA. When comparing with the normal model using SIMCA classification of the autofluorescence spectra after 337-nm excitation, 10 of 14 normal samples were classified as members of the normal group, whereas all 12 lesion samples were found to be nonmembers. When comparing with the lesion model, 8 of 12 lesions were classified as members of the lesion group, whereas only 9 of 14 normal samples were found to be nonmembers. In all, four normal samples were classified as belonging to both groups, and three normal and four lesions as belonging to neither of the groups. If we define sensitivity as the proportion of lesions classified as belonging to the lesion group only, it will be 8 of 12 (67%). Accordingly, defining the specificity as the proportion of normal samples classified as belonging to the normal group only will give a value of 6 of 14 (43%). The total success rate, hence, would be 14 of 26 (54%).

PCA and logistic regression. Four principal components explain 98.8% of the spectral variance using a PCA calibration of all autofluorescence spectra from 337-nm excitation (Fig. 2a). The first principal component explains
the largest variations within the data set, the second component is the second most important, etc. The resulting scores were used in equation (2) to calculate the posterior probabilities that the samples belong to the normal group. The cost of misclassifying a normal sample was chosen to 0.25 for maximal number of correctly classified samples. Twenty-one of 25 lesions had a probability of less than 50% of being normal, and 26 of 29 normal samples had a probability of more than 50%. Hence, the sensitivity and specificity were 84% and 90%, respectively, and the total success rate was 87%.

**PLS and logistic regression.** The same logistic regression algorithm was applied to scores from a PLS model of the same spectra (autofluorescence from 337-nm excitation). Four PLS components were used, giving an RMSEP of 0.70, and explaining 99% and 53% of the variance in X and y, respectively. The four PLS component spectra are shown in Figure 2b. It can be seen that the important variations are in the lower wavelength regions, close to the fluorescence peaks of collagen and NADH. The scores from this calibration were used in the logistic regression algorithm (equation 2). A cost of misclassification of normal as lesion of 0.35 was optimal. The result can be seen in Figure 3. Twenty-two of 25 lesions had a posterior probability of being normal less than 50%, whereas 27 of 29 normal spots had a probability of more than 50%. Thus, the sensitivity was 88%, the specificity 93%, and the total success rate 91%. All three misclassified lesions were dysplasias.

**PLS-DA.** PLS discriminant analysis was also performed, using the PLS calibration. Now the y values were predicted directly by the PLS model. Full cross-validation was used. Ideally, all normal values would have a predicted y value of +1 and all lesions a value of −1. For the autofluorescence spectra from 337-nm excitation, the optimal threshold was determined as described above to be 0.10, giving the result shown in Figure 4. Twenty-two of 25 lesions and 27 of 29 normal tissue spots were correctly classified, giving a sensitivity of 88%, a specificity of 93%, and a total success rate of 91%. This result is identical with the one obtained using PLS together with logistic regression. All but one of the spectra that failed classification were the same in both cases.

The results from all classification algorithms for autofluorescence spectra from 337-nm excitation are shown in Table 1. PCA with logistic regression, PLS with logistic
regression and PLS discriminant analysis all give comparable results. Three dysplastic lesions and one normal tissue spot were misclassified using both PCA+logistic regression, PLS‡logistic regression and PLS-DA.

**Fluorescence From 405-nm Excitation Using ALA**

The PLS-DA algorithm was now used as above on fluorescence spectra from 405-nm excitation in patients given ALA before the examination. These spectra range from 463 to 650 nm with a wavelength resolution of 3 nm. The normalized spectra from normal tissue and lesions can be seen in Figure 5. On an average, the lesion spectra show a more pronounced PpIX fluorescence at 635 nm and a lower autofluorescence compared with the normal spectra.

A PLS model was built with the optimal prediction using three PLS components (RMSEP = 0.77). Three components explain 99% of the variance in X and 45% of the variance in y. The PLS component spectra are shown in Figure 6. The PpIX peak is an important feature in all three components, and in the second and third the autofluorescence also comes in.

Using the model for prediction gives the result shown in Figure 7. The threshold, chosen from the criteria described above, was set to 0.45, giving a sensitivity of 100% (13 of 13) and a specificity of 69% (18 of 26) for this data set.

**DISCUSSION**

The investigation showed that multivariate classification algorithms applied to fluorescence emission spectra can be used in vivo to discriminate laryngeal lesions from normal tissue. Apart from SIMCA, all algorithms evaluated in this study classify lesions from normal tissue with a success rate of more than 85%. SIMCA classification performed worse than the other three algorithms, as can be seen in Table 1. One problem here might be that dividing the data in separate calibration and test sets resulted in too small calibration sets. Only 15 and 13 spectra were used to determine the normal and lesion PCA model, respectively. In addition, the SIMCA classification algorithm does not assume two mutually exclusive groups of data, but tested spectra can be classified as belonging to both groups or neither. Talking about sensitivity and specificity in this case may be dubious, but we give them the strictest possible definitions, including only the spectra that are classified as belonging to one group only. Hence, the SIMCA results presented in Table 1 are not fully comparable to the others and have to be interpreted with caution.

Looking at the other three methods, the results are comparable. The intention was not to determine which algorithm is superior, but rather to study their possibilities. Using PCA and logistic regression with the four first principal components works well, but substituting the principal components with PLS components gives a somewhat better result on this set of data. Using a logistic regression algorithm has the advantage of providing posterior probabilities, which might be more easily interpreted clinically. In addition, it seems to separate the

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SIMCA (%)</th>
<th>PCA + log. regr. (%)</th>
<th>PLS + log. regr. (%)</th>
<th>PLS-DA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>8/12 (67)</td>
<td>21/25 (84)</td>
<td>22/25 (88)</td>
<td>22/25 (88)</td>
</tr>
<tr>
<td>Specificity</td>
<td>6/14 (43)</td>
<td>26/29 (90)</td>
<td>27/29 (93)</td>
<td>27/29 (93)</td>
</tr>
<tr>
<td>Total success rate</td>
<td>14/26 (54)</td>
<td>47/54 (87)</td>
<td>49/54 (91)</td>
<td>49/54 (91)</td>
</tr>
</tbody>
</table>

*SIMCA, soft independent modeling of class analogy; PCA, principal component analysis; log. regr., logistic regression; PLS-DA, partial least squares-discriminant analysis.
groups better than PLS-DA, making the classification result less sensitive to the choice of threshold. However, the logistic analysis step does add extra complexity to the algorithm, because a normal probability density curve has to be fit to each principal component and each group. PLS-DA is much simpler and seems to classify the samples equally well, but it yields a predicted \( y \) value instead of the more easily interpreted posterior probability for each measurement. PLS-DA also performed well on fluorescence spectra from patients who had been given ALA as a photosensitizer before the investigation. Sensitivity and specificity were good both looking at autofluorescence after 337-nm excitation and using 405-nm and ALA as a photosensitizer. The results suggest that ALA might not add much information to the classification, due to the strongly varying amounts of PpIX both in normal and lesion tissue.

In this study, we have only included fully normal tissue and premalignant and malignant lesions. Areas of inflammation and other conditions of the tissue that did not fit into either of these two groups were excluded from the analysis. Including more types of tissue in the classification would result in a more complex analysis, and possibly in a lower sensitivity and specificity. If a reliable model is available, based on a relevant population, these results suggest that fluorescence measurements together with multivariate analysis in principle could be used for obtaining a real time diagnosis of the tissue during examination.

In summary, our in vivo point measurements suggest that fluorescence spectroscopy might be a useful technique to discriminate between normal tissue and lesions in the laryngeal region. Using multivariate analysis, such as PLS-DA or PLS in combination with logistic regression, can yield good diagnostic potential. This could be useful to assist the examining doctor in distinguishing lesion borders during surgical removal of lesions and as an aid guiding the endoscopist to the sites most relevant for biopsy sampling.

ACKNOWLEDGMENTS

The authors thank Dr. Martin Andersson, Department of Technical Analytical Chemistry, Lund Institute of Technology and Dr. Andrew Berger, Beckman Laser Institute and Medical Clinic, University of California, Irvine for their useful discussions. We also thank Prof. Sune Svanberg, Department of Physics, Lund Institute of Technology for support.

REFERENCES

11. Kleemann D, MacRobert AJ, Mentzel T, Speights PM, Bown SG. Photodynamic therapy on the normal rabbit larynx with...


