

Characterization of healthy skin using near infrared spectroscopy and skin impedance

Ida Bodén · David Nilsson · Peter Naredi ·
Britta Lindholm-Sethson

Received: 29 June 2007 / Accepted: 26 March 2008
© International Federation for Medical and Biological Engineering 2008

Abstract Near infrared spectroscopy (NIR) and skin impedance (IMP) spectroscopy are two methods suggested for diagnoses of diseases inducing adverse effects in skin. The reproducibility of these methods and their potential value in non-invasive diagnostics were investigated. Measurements were performed in vivo on healthy skin at five anatomic body sites on eight young women. partial least squares discriminant analysis showed that both methods were useful for classification of the skin characteristics at the sites. *Inter-individually* the NIR model gave 100% correct classification while the IMP model provided 92%. *Intra-individually* the NIR model gave 88% correct classification whereas the IMP model did not provide any useful classification. The correct classification was increased to 93% when both datasets were combined, which demonstrates the value of adding information. Partial least squares discriminant analysis gave 72% correct predictions of skin sites while the combined model slightly improved to 73%.

Keywords Near infrared diffuse reflectance spectroscopy · Skin impedance · Multivariate data analysis · Skin characterization · Reproducibility

1 Introduction

Many diseases are reflected in abnormal changes in the skin. Therefore, there is an increasing interest in the development of methods for diagnosing various illnesses by means of non-invasive skin measurements. Near infrared (NIR) spectroscopy and skin impedance (IMP) are two non-invasive skin characterization techniques suggested for the diagnosis and monitoring of various markers of diseases, such as neuropathy, blood glucose levels and microcirculation in patients with diabetes [4, 5, 9, 12, 19, 25, 28]. The two techniques are also suggested for monitoring radiotherapy induced erythema [18].

Deviations in the optical properties of human skin in vitro between different individuals and different skin characteristics from eyeballing NIR spectra were reported to be fairly small by Troy and Thennadil [23]. Surprisingly, no attempt was made in that study to use multivariate analysis of the data, which has proven to be an efficient approach. Although it is possible to select a few variables from the achieved NIR data and hope these would represent some important factors that best define the skin, this approach would be too risky and a lot of information would get lost. Therefore, data reduction methods using projection, such as principal component analysis (PCA) [26] and partial least squares discriminant analysis (PLS-DA) [8, 21, 22, 27], were chosen in our investigation to enable the identification of subtle differences in the skin. Other possible options could have been genetic algorithms or neural

I. Bodén · P. Naredi
Department of Surgical and Perioperative Sciences/Surgery,
Umeå University, 901 87 Umeå, Sweden

I. Bodén · B. Lindholm-Sethson
Department of Chemistry, Umeå University,
901 87 Umeå, Sweden

I. Bodén · B. Lindholm-Sethson (✉)
Centre for Biomedical Engineering and Physics,
Umeå University, 901 87 Umeå, Sweden
e-mail: Britta.sethson@chem.umu.se

D. Nilsson
UmBio AB, Box 7978, 907 19 Umeå, Sweden

networks, but these methods do not provide statistical significance to the outcome. In an earlier study [20], multivariate statistic is used to simplify complex data for univariate statistical analyses. In a similar way scores from PCA models have been used, in our investigation, to define a *scatter value* (SV) as a measure of the scatter of measurements in a PCA score plot. These SVs have then been analysed univariately with the Kruskal–Wallis test [14].

The present paper highlights the fact that NIR and IMP measurements indeed provide valuable information on skin characteristics, *in vivo*. This is not surprising, since skin characteristics, such as moisture [17, 24] and skin thickness [15], vary depending on anatomical body site, age, sex and sun exposure. In adults, the thickness of the skin varies from a few millimetres at the eyelid up to a centimetre at the foot sole [15]. Most present methods for evaluating skin and diagnosing skin diseases are subjective and some are time consuming, expensive and also invasive. Improved non-invasive and objective diagnostic methods would be valuable both from an economic point of view and with regard to quality in health care.

The concepts *sensitivity* and *specificity* are used to compare the quality of various diagnostic methods. Sensitivity is the proportion of reference test positive subjects who test positive while specificity is defined as the proportion of reference test negative subjects who test negative. A relevant non-invasive method should have 100% sensitivity, i.e., all positive subjects must be found. The ABCDE method [7] is a simple and objective tool for the preliminary diagnosis of cutaneous melanomas where the physician evaluates the following parameters: area, border, colour, diameter and erosion of the nevi. However, the sensitivity in this method for clinical diagnosis of malignancy is 91.3% [11]. This relatively low sensitivity leads to a “better safe than sorry” mentality and many benign nevi are excised unnecessarily. The diagnostic accuracy of the ABCDE method can be improved by 50% if using a dermatoscope, on condition that the examination is performed by an expert [13]. Aberg et al. [1] showed that it was possible to use either a micro-invasive or a non-invasive skin IMP probe to classify basal cell carcinoma (BCC) and malignant melanoma (MM) with 96% sensitivity and 86% specificity for BCC and 92% sensitivity and 80% specificity for MM. These measurements were analysed using PCA and Fisher’s discriminant analysis [1]. Har-Shai et al. [10] reported a sensitivity of 91% and a specificity of 64% in an attempt to diagnose melanoma using a micro-invasive probe in IMP. When the micro-invasive IMP measurements were combined with image analysis, the sensitivity for melanomas was increased to 100% with only a slight decrease in specificity to 62% [10]. In this later study the IMP measurements were only analysed by comparing the conductivity, at 2 kHz, for a

suspected mole with the conductivity for a benign nevus. It is also unclear how the data from the two techniques were combined. However, these findings still indicate that IMP is a useful tool for diagnosing skin diseases, such as malignant melanoma, especially combined with other techniques.

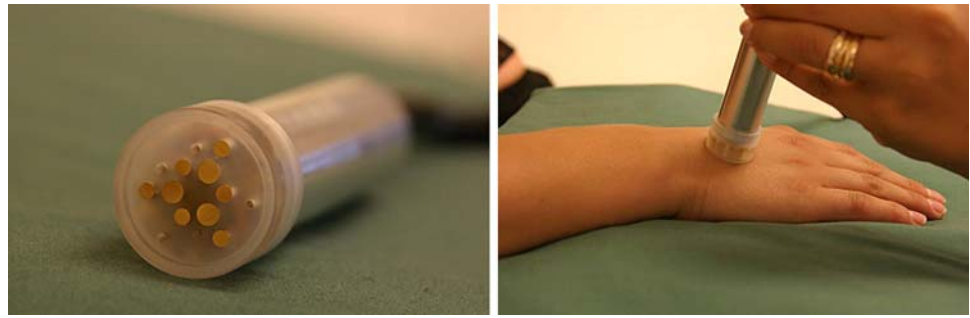
The aims of the present study on healthy individuals are, first, to investigate and evaluate the reproducibility of measurements using skin IMP and NIR spectroscopy; second, to investigate if the two methods can differentiate between skin at various anatomic body sites with different characteristics; and third, to demonstrate the added value of combining data collected using the two techniques. We have indeed seen that the measurement methods used were reproducible and robust. Both NIR spectroscopy and skin IMP were useful methods to differentiate between skin measurements at different anatomic sites and the classification of the skin characteristics was improved when adding the information from the NIR spectra and the IMP measurements together.

2 Materials and methods

2.1 Instrumentation

A Bruker MatrixF FT-NIR instrument (Bruker Optics, Täby, Sweden) equipped with a fibre-optic reflectance probe was used for NIR measurements. Diffuse reflectance spectra between 12,500–4,550 cm^{-1} (800–2,200 nm) with a resolution of 4 cm^{-1} were collected and displayed by OPUS software (Bruker Optics, Täby, Sweden). The raw data was then transferred to Microsoft Excel software and Matlab (The MathWorks Inc., Kista, Sweden) for further analysis. The optical radiated power emitted from the instrument is 21.8 mW. After loss in the fibre and in junctions a value of emitted power from the probe is 1.20 mW (Bruker Optics, Täby, Sweden). The skin IMP measurements were performed with a home-built PC-based instrument. A 16 MHz function generator, NI PCI-5401 generates an AC current that is injected into the skin and the AC voltage that is generated is picked up by a high-speed two channel PC-based oscilloscope, NI 5112. A program written in LabVIEW (National Instruments Sweden AB, Solna, Sweden) was used to control the generation and acquisition of signals. The IMP measurements were accomplished with flat circular gold electrodes arranged on a plane in a hexagonal pattern and at 51 frequencies logarithmically distributed between 100 Hz and 10 MHz measurements (Fig. 1). A vacuum pump was applied to the probe to maintain constant pressure between the probe and the skin. The apparatus was tested for patient safety and approved by the department of Medical Technology

Fig. 1 The impedance probe. A hexagonal pattern of the gold plates was used in this study



and Informatics (MTI) at Umeå University hospital, for use in clinical investigations.

2.2 Skin measurement procedure

Eight young women, aged 15–19 years volunteered to participate in the study. NIR spectra and skin IMP measurements were obtained at five different anatomic body sites, i.e., back of hand (A), inner upper arm (B), back (C), calf (D) and cheek (E). In this study, the skin at each anatomic body site represents specific skin characteristics described with for instance elasticity, thickness and skin touch. The skin was prepared for measurements in different conditions according to Table 1. Five replicates were measured for each skin site and condition. This gives a total number of 80 measurements on the same anatomic body site.

2.3 Multivariate data analysis

PCA [26] and PLS-DA [21, 22] were performed with EVINCE (UmBio AB, Umeå, Sweden). All matrices were variable mean centred but not scaled during the analysis.

MATLAB was employed for mathematic editing of the matrixes and to evaluate statistical differences with the Kruskal–Wallis test, where $\alpha = 0.05$ [14].

2.4 Pre-treatment of data

The skin IMP measurements yielded real and imaginary values at 51 frequencies and from NIR measurements

pseudo-absorbance values were obtained at 2,281 wavelengths.

All NIR data were combined in the raw NIR matrix (1,000, 2,281) where each row corresponds to one measurement consisting of 2,281 absorbance values. Outliers were identified by falling outside a 95% Hotelling T^2 area in the score plot and/or by extensively exceeding DCrit, the critical distance, (95%) in a DModX plot, which shows the distances from the observations to the model [6]. The raw NIR matrix was analysed with PCA and one observation was identified as an outlier and deleted from the matrix (Table 2).

The absolute IMP, IZI and phase angle Φ , were transformed into complex IMP, i.e., a real part and an imaginary part from the following equation:

$$Z = |Z|e^{j\phi} = \text{Re } Z + j\text{Im } Z = R + jX. \tag{1}$$

Thus, the complex IMP could be calculated for every frequency:

$$\text{Re } Z = |Z| \cos(\Phi); \text{Im } Z = |Z| \sin(\Phi). \tag{2}$$

The raw IMP matrix (Table 2) was constructed where each row corresponds to a unique measurement. The imaginary IMP values are found in the first 51 columns and the real IMP values follow. PCA was performed on this matrix and a score plot of the first two principal components indicated large individual differences. In the attempt of diminishing this effect, an individual mean spectrum was calculated from the average of all spectra of each person. Each IMP spectrum was normalized with the corresponding mean IMP spectrum by means of point-by-point division. Thus,

Table 1 An overview of the various pre-treatments of the skin before measurements

	Skin site				
	A (back of hand)	B (inside of upper arm)	C (back)	D (calf)	E (cheek)
IMP					
Right side	S, L	S, L	S, L	S, L	S, L
Left side	S	S	S	S	S
NIR					
Right side	NO, S, L	NO, S, L	NO, S, L	NO, S, L	NO, S, L
Left side	NO, S	NO, S	NO, S	NO, S	NO, S

S Soaking with saline solution for 90 s before measurements, L Treatment of measurements site with lotion, NO Neither soaking nor lotion

Table 2 Matrices and multivariate models

Matrix	<i>N</i>	<i>K</i>	PCA model			PLS-DA model			
			# excluded outliers	# comp	R ² X (1st: 2nd: 3rd component)	# excluded outliers	# comp	R ² Ycum	Q ² Ycum
1 Raw NIR matrix	1,000	2,281	1 ^a	25	0.56: 0.19: 0.11	–	–	–	–
2:1 Raw IMP matrix	600	102	3	4	0.89: 0.08: 0.02	–	–	–	–
2:2 IMPcorr matrix	600	102	3	4	0.78: 0.14: 0.03	–	–	–	–
3 IMP data set	600	102	2	10	0.76: 0.15: 0.03	–	–	–	–
4 NIR data set	599	2,281	0	7	0.53: 0.21: 0.11	–	–	–	–
5 NIRIMP data set	599	2,383	2	9	0.36: 0.28: 0.11	–	–	–	–
6 IMPred data set	199	102	3	8	0.75: 0.14: 0.04	3	7	0.54	0.27
7 NIRred data set	199	2,281	3	20	0.56: 0.21: 0.09	3	5	0.50	0.46
8 NIRIMPred data set	199	2,383	3	21	0.40: 0.25: 0.11	3	8	0.57	0.48
9 NIRmIMPm data set	80	2,383	–	–	–	0	4	0.43	0.36
10 NIRm data set	80	2,281	–	–	–	0	5	0.51	0.43
11 IMP data set	80	2,383	–	–	–	0	4	0.36	0.25
PLS-DA model for predictability	<i>N</i>	<i>K</i>	# objects excluded and reused as test set			# comp	R ² Ycum	Q ² Ycum	
Based on matrix 7. NIRred data set									
AN	196	2,281	25			5	0.62	0.57	
CE	196	2,281	24			6	0.66	0.59	
IS	196	2,281	24			5	0.62	0.57	
KA	196	2,281	24			5	0.64	0.61	
KR	196	2,281	24			5	0.61	0.56	
MA	196	2,281	25			5	0.63	0.59	
ME	196	2,281	25			5	0.63	0.57	
SA	196	2,281	25			5	0.63	0.59	
Based on matrix 8. NIRIMPred data set									
AN	196	2,383	25			8	0.57	0.47	
CE	196	2,383	24			8	0.73	0.65	
IS	196	2,383	24			10	0.77	0.61	
KA	196	2,383	24			8	0.74	0.64	
KR	196	2,383	24			8	0.71	0.60	
MA	196	2,383	25			8	0.73	0.62	
ME	196	2,383	25			9	0.75	0.62	
SA	196	2,383	25			8	0.72	0.63	

^a The outlier was due to failure of measurement and, therefore, deleted from all the other matrices were NIR spectroscopy data was used

each row in the new matrix became a unit-less vector less dependent on the individual. This new matrix was named: IMPcorr matrix.

One important aim of the present study was to demonstrate the added value of the information from NIR spectroscopy and skin IMP. Since there were no corresponding IMP measurements, the measurements on dry skin with NIR spectroscopy were removed from the raw NIR matrix. The two matrices, the IMPcorr matrix and the reduced raw NIR matrix, were scaled into same variance, named NIR data set and IMP data set, and combined to one

NIRIMP data set consisting of 599 observations and 2,383 variables (Table 2).

The three obtained data sets were reduced to contain only objects measured on soaked skin at the right side of the body. Unless otherwise mentioned, all the following analyses were based on these three reduced data sets, i.e., the IMPred data set, the NIRred data set and the NIRIMPred data set (Table 2).

Three outliers were identified and excluded from a PCA model based on the combined NIRIMPred data set described in the previous section. The same outliers as

above were also removed from the IMPred data set and the NIRred data set and different PCA models were generated on these reduced matrixes as described in the result section (Table 2).

PLS-DA was performed on the NIRIMPred data set and three outliers were identified and removed. These outliers were excluded also from the IMPred data set and the NIRred data set and different PLS-DA models were generated on these reduced matrixes (Table 2).

PLS-DA was performed on the NIRred and NIRIMPred data sets in order to determine the predictability. The measurements from the different individuals were used as test sets; i.e., a total of eight PLS-DAs for each data set were created (Table 2). The test sets were then predicted back into the models and the number of mispredicted samples was recorded.

From the NIRIMP data set, an average of the vectors of the replicates in each group of measurements was calculated. However, measurements on lotion treated skins were excluded. This was also done with the NIR and IMP data sets resulting in the NIRmIMPm, NIRm and IMPm data sets. PLS-DA models were created as above (Table 2).

2.5 Reproducibility

The PCA three-dimensional score plots based on the IMPred data set and the NIRred data set were used to find a measure of the reproducibility of the measurements. The aspects of reproducibility were considered as described below and also illustrated in Fig. 2. The five replicates measured at each skin site form a group and a centre of gravity was calculated for each such group. This gave five groups for each individual and totally 40 groups. The Euclidian distances between the centre of gravity and the five scores within each group was then calculated and as a measure of deviation SV was calculated (Fig. 2 and Table 3). The Euclidian distance was chosen as a measure of reproducibility instead of the more scale independent Mahalanobis distance, which is related to leverage and outlier detection [16].

The intra-individually calculated SVs were used to estimate the reproducibility of NIR and IMP measurements on different skin sites with different characteristics within the same model. Inter-individual reproducibility aspects were evaluated by considering all scores from measurements at one skin site as a group. In this case only five groups with 5×8 scores in each were prevalent. Five inter-individual SVs were calculated by the same method as already described and defined as “ALL” in Table 3. These SVs were used to compare inter- and intra-individual reproducibility.

Inter-individual reproducibility aspects were also evaluated by calculating the *mean SV* (Table 3). The mean SV was defined as the average SV \pm standard deviation of a

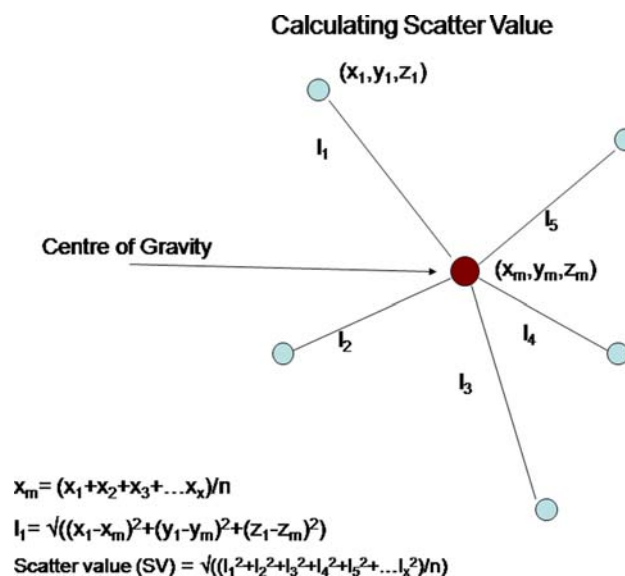


Fig. 2 The figure shows a visualization of the calculation of the scatter values (SV). The y_m and z_m were calculated by the same formula as for x_m . Since the centre of gravity is a middle point and centralize the observation points its coordinates become (0, 0, 0). The SV is a measure of deviation of the distances between the centre of gravity and the observation points

specific set of SVs, for example: the SVs of different skin sites originating from one person, or the SVs of one specific skin site originating from all persons. The mean SVs were derived to get an idea of the size of the scattering in a particular set. However, in order to investigate if there was a significant difference ($p < 0.05$) between the sets, the medians of each set were compared using the non-parametric Kruskal–Wallis test [14]. The Kruskal–Wallis test was thus performed with respect to both person and skin site for each measurement technique, i.e., NIR and IMP.

The *total mean SV* is defined as the average SV \pm standard deviation of all the SVs originating from all measurements with NIR and IMP, respectively. The relative standard deviation (RSD) was calculated for each method and used to compare the reproducibility of two measurements methods, i.e., NIR spectroscopy and skin IMP (Table 3).

3 Results

Firstly, selected score plots from PCA were inspected for a first overview of data and also to obtain some guidelines for further analysis of it. Then we examined the SVs in order to judge the reproducibility of the measurement methods. The possibility for classification with PLS-DA is examined based on both intra- and inter-individual approaches on different skin sites. Finally, the PLS-DA was done to investigate the prediction ability in our data.

Table 3 Scatter values

	AN	CE	IS	KA	KR	MA	ME	SA	ALL	Mean SV ± SD
SVs based on score values from NIRred PCA model										
A	0.61	0.58	0.82	0.95	0.77	0.60	1.11	0.70	1.82	0.77 ± 0.39
B	1.18	1.25	0.67	0.91	1.76	1.54	1.10	1.01	1.82	1.18 ± 0.39
C	0.71	0.62	0.96	0.57	0.76	0.78	0.84	1.29	1.78	0.82 ± 0.38
D	0.80	0.95	0.90	1.02	1.13	0.97	0.69	0.55	2.13	0.88 ± 0.45
E	1.10	0.88	1.20	0.64	1.22	0.86	2.20	1.72	2.93	1.23 ± 0.74
Mean SV ± SD	0.88 ± 0.25	0.86 ± 0.27	0.91 ± 0.20	0.82 ± 0.20	1.13 ± 0.41	0.95 ± 0.36	1.19 ± 0.36	1.05 ± 0.59	2.09 ± 0.47	0.97 ± 0.36
Total SV ± SD										0.37
RSD										
SVs based on score values from IMPred PCA model										
A	2.22	0.41	1.08	1.81	1.26	1.33	3.23	0.92	2.19	1.53 ± 0.85
B	0.58	0.91	1.41	1.15	0.53	0.70	0.78	0.55	2.01	0.83 ± 0.49
C	1.18	1.03	1.32	2.35	1.06	0.72	1.57	0.82	2.09	1.26 ± 0.56
D	0.88	0.86	0.67	0.83	1.17	0.82	0.58	0.36	1.86	0.77 ± 0.43
E	2.10	1.16	0.70	0.84	0.85	0.95	0.91	0.51	1.93	1.00 ± 0.55
Mean SV ± SD	1.39 ± 0.73	0.87 ± 0.28	1.04 ± 0.34	1.40 ± 0.67	0.97 ± 0.29	0.90 ± 0.26	1.4 ± 1.1	0.63 ± 0.23	2.01 ± 0.13	1.08 ± 0.58
Total SV ± SD										0.54
RSD										

3.1 Principal component analysis

A slight tendency to separation between measurements from the right and left side of the body was observed in the PCA score plot from the NIR data (data not shown). This might be due to artefacts such as the different bending of the fibre-optic bundle in the NIR probe or physiological differences between the skin on the left and the right side of the body. The reason for this minor difference is not yet fully understood and a future investigation is planned. To avoid introduction of systematic errors only measurements from the right side of the body were employed in the following unless otherwise mentioned. Measurements on lotion treated sites showed a tendency to group in the PCA IMP model and were shifted intra-individually in the PCA NIR model.

Soaking the skin with saline solution is essential for reproducible IMP measurements [3]. This decreases the impedance of the stratum corneum and ensures skin penetration. For the NIR measurements in this study, soaking the skin caused a minor shift intra-individually, but not inter-individually, in a PCA plot on the reduced NIR raw matrix (Table 2) compared to dry measurements. However, the main features of the plot did not change.

3.2 Reproducibility

The reproducibility of the measurements was assessed by the *scatter value* (SV) for each skin site (Table 3). The lower the SV, the tighter was the clustering of the scores and the reproducibility was, therefore, classified as high and vice versa. The first three components in each of the models explained the variation to different degrees. Therefore, SVs can only give a relative measure of reproducibility and we can only compare the scatter within each model.

3.2.1 Influence from skin characteristics on scatter values

The 40 *intra-individual* SVs from either NIR or IMP measurements showed some variation with no clear pattern for all five skin sites within each individual person. No obvious relationship could be detected between the highest and lowest intra-individual SVs from the five skin sites and the highest and lowest inter-individual SVs, neither in NIR nor in IMP.

The *inter-individual* SVs for the five different skin sites from all eight women were calculated for the two methods separately. From the NIR data, it was observed that the highest value was obtained for the cheek (skin site E) and the lowest for the back (skin site C), i.e., 2.93 and 1.78. In the IMP measurements the back of hand (skin site A) showed the highest SV and the calf (skin site D) the lowest with 2.19 and 1.86, respectively.

To investigate if there is a *general* difference in reproducibility between measurements at different skin sites, the medians of SVs for the five skin sites were analysed with the Kruskal–Wallis test. In the NIR model, the highest mean SV was 1.23 ± 0.74 (skin site E) and the lowest mean SV was 0.77 ± 0.39 (skin site A) and a significant difference ($p < 0.05$) between the reproducibility of the different skin sites was found. In the IMP model, the highest mean SV was 1.53 ± 0.85 (skin site A), the lowest was 0.77 ± 0.43 (skin site D), and also in this case a statistically significant difference ($p < 0.05$) in reproducibility was obtained when comparing measurements on different skin sites to one another.

3.2.2 Influence from individual variations on scatter values

A mean SV for each person was calculated from the SVs for all individual skin sites for each method separately. The highest mean SVs from the NIR model was 1.19 ± 0.59 (Individual: ME) and the lowest 0.82 ± 0.20 (Individual: KA) whereas for the IMP model the highest and lowest mean SV were 1.4 ± 1.1 (Individual: ME) and 0.63 ± 0.23 (Individual: SA) for all the skin sites from each person, respectively. A Kruskal–Wallis test indicated no significant difference ($p < 0.05$) between medians of the different groups in neither the NIR nor the IMP model. This implied that the reproducibility in the two measurement methods do not vary in relation to which person was measured.

3.2.3 Influence of measurement method on scatter values

The total mean SV were calculated from both NIR and IMP measurements for all 40 SVs and the absolute values found were 0.97 ± 0.36 and 1.08 ± 0.58 , respectively. The relative standard deviations (RDS) were 0.37 for NIR and 0.54 for IMP measurements.

3.3 Classification with simple planar discriminant analysis in PLS-DA score plots

Here, we wanted to investigate whether the addition of data from two measurement methods improved the classification

ability. Would the overlap decrease between scores from measurements on different skin sites, i.e., different skin characteristics? In this analysis PLS-DA were performed on the NIRred data set, the IMPred data set and the NIRIMPred data (Table 2 and Fig. 3). A visual inspection of the PLS-DA IMPred model showed there was a large overlap between groups of different skin characteristics. This was in contrary to the PLS-DA NIRred model where the overlap was minor. The PLS-DA NIRIMPred model appeared to give an even smaller overlap than the NIRred model, which was confirmed in the detailed analysis below.

A simple manual planar discriminant analysis was performed in the following manner. A plane was drawn in the three-dimensional PLS-DA score plots between groups of measurements on skin from two different anatomic body sites at a time. A total of ten planes were drawn and in six of the cases no overlapping of the groups were seen in neither of the models. In cases with overlap, the plane was tilted to obtain an equal number of scores with wrong classification at each side of the plane.

In the four remaining cases 2 + 2 (skin sites D and E), 3 + 3 (skin sites C and D), 4 + 4 (skin sites B and D), and 5 + 5 (skin sites C and E) (total 28) objects were misclassified in the NIRred model whereas in the NIRIMPred model only 1 + 1 (skin sites C and E), 2 + 2 (skin sites C and D), 4 + 4 (skin sites B and D) and 4 + 4 (skin sites D and E) (total of 22) were misclassified. Thus, NIR data provided 86% correct classification but when combining the NIR and the skin IMP data the correct classification was increased to 89%.

The resulting PLS-DA three-dimensional score plots of the *NIRm data set*, *IMPm data set*, and the *NIRmIMPm data set* are shown in Fig. 4a–c. Ten planes were again drawn in the NIRm and NIRmIMPm models and the manual planar discriminant analysis was repeated. Six of the cases showed no overlap for the NIRm model and seven of the cases for the NIRmIMPm model. Thus when an average of the replicates was used the correct classification improved from 83% (total 14 misclassified out of 80) for the NIRm data to 90% correct classification (total 8 misclassified out of 80) for the combined NIRmIMPm data set (Fig. 4b, c).

Fig. 3 Three-dimensional score plot of PLS-DA models of **a** IMP model, **b** NIR model and **c** the combined data NIRIMP model from soaked samples on right side of body. Skin site A: back of hand (green), B: inner upper arm (dark blue), C: back (yellow), D: calf (light blue), E: cheek (purple)

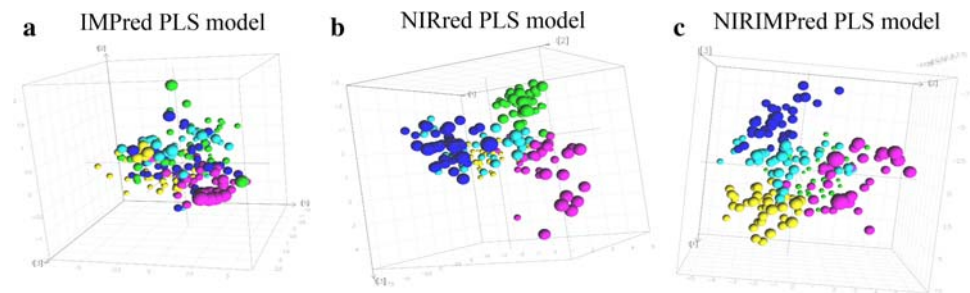
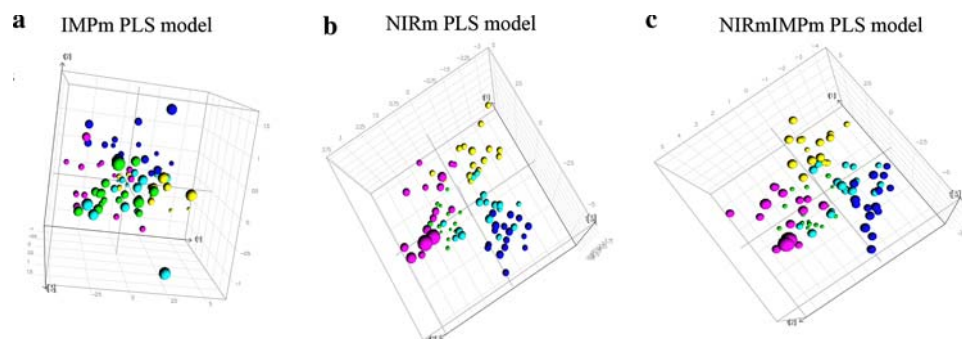


Fig. 4 Three-dimensional score plot of PLS-DA models from an average of the replicates in the **a** IMP model, **b** NIR model, **c** the combined NIRIMP model. The figures show the separation between the different skin sites for each model. Skin site A: back of hand (green), B: inner upper arm (dark blue), C: back (yellow), D: calf (light blue), E: cheek (purple)



3.4 Prediction with PLS-DA

To judge the predictability in the NIR model and the NIRIMP model a PLS-DA was performed and the numbers of correctly predicted samples were recorded. The number of correct predicted samples was 142 out of 197 (72.1%) for the PLS-DA NIR model and 143 out of 197 (72.6%) for the PLS-DA NIRIMP model. The major observed change was that in the NIR model 2 out of 39 (5%) of the skin site D samples were correctly predicted whereas in the NIRIMP model this was improved to 11 out of 39 (29%). The predictability of the other skin sites were either improved or decreased slightly when comparing the NIR model to the NIRIMP model.

4 Discussion

We have measured NIR spectroscopy and skin impedance on five anatomic body sites on eight healthy young women in a tight age group. The intra-individual differences of the skin were larger than the inter-individual differences. Each measurement at a specific skin site, including replicates, can, therefore, be considered as one independent measurement. Thus, there are 80 independent measurements, which is possible only since we have a homogeneous group of persons.

In NIR spectroscopy the dominating feature is the water peak and, therefore, the technique is known to be very much dependent on the water content in the measured sample. It is shown by others that the intensity and shapes of the water bands change with skin hydration [17, 24]. Surprisingly, soaking the skin with a saline solution did not lead to any large effects in the PCA score plot from the NIR model. Only a small shift was observed in the scores from the soaked sites compared to those from dry sites and, therefore, results from measurements on soaked skin could be used with confidence. The observed small effect from soaking the skin in the score plot of NIR data might be explained with the homogenous data set with eight young healthy women. The moisture content in the skin of a more

heterogeneous data set with respect to age and gender is expected to vary considerably [17, 24]. Soaking the skin before measurements might reduce this expected variation in data and will be explored in a future study. This is of great importance since IMP measurements are not at all feasible on dry skin and in a newly developed combination probe both measurements methods will be performed almost simultaneously in the same plane of the probe head (Britta Lindholm-Sethson and Paul Geladi, 2001, “Device for non-destructive and non-invasive clinical measurements”, Patent pending, EPO nr 027323201.5).

Moreover, in the scores from measurements on lotion treated sites a small overall shift was observed in the PCA score plot from the NIR model whereas the effect was larger in the IMP model. Thus, when visually studying the PCA score plots of the raw IMP data set the biggest effect was from the lotion. This suggests that patients and test persons should be informed not to treat the skin with lotion or makeup the same day as the measurement.

Another important observation was that no block effect was seen between the six first measured persons and the last two that were measured 1 year later (data not shown). The instruments were frequently used in other project during that period of time and this showed that the methods indeed are robust and reliable also over time.

4.1 Reproducibility

A prerequisite for predictive measurements with a NIR-IMP-probe is reproducible data. Therefore, one main emphasis in the present investigation is to give a practical definition on reproducibility for our purpose. Thus we have introduced the scatter values, SV.

It could be assumed that the variation between individuals would be larger than the inter-individual variation between skin sites. However, this was not the case in neither of the two models where a statistically significant difference between the medians of the SVs of different skin sites was found with the Kruskal–Wallis test, whereas no significant difference was observed when comparing different persons. This might be true only for the very

homogeneous data set in this study. In a more heterogeneous data set, larger individual differences might appear which could be corrected in a similar way as the IMP data set in the present paper.

The precision in measurements varied between the skin sites in both methods intra-individually according to the Kruskal–Wallis test. Therefore, the prediction ability will probably vary depending on the skin site of the measurements. No general conclusions can be drawn regarding a less homogeneous data set. A larger disparity in reproducibility would likely have been found in an investigation with measurements on both sexes and a variability in age, ethnicity and health conditions. However, the present investigation provides the basis for future studies on a more heterogeneous material.

The *total mean SVs* from NIR and IMP showed that the precision in IMP measurements was marginally lower than in the NIR measurements and, therefore, the reproducibility in the two methods might differ somewhat. The large overlap between groups of measurements on different body anatomic sites in the IMP models compared to the NIR model (Fig. 3a, b) cannot be explained only with the slightly less reproducible method. The information on skin characteristics is probably more diverse in NIR spectra than in the results from IMP measurements.

4.2 Classification

From the PLS-DA models it was visualised that there were information on skin characteristics in both measurement methods (Figs. 3, 4) since measurements from different skin sites were totally or partially separated. This was particularly obvious in the PLS-DA score plots from each individual where the correct classification was 100% for the NIR and NIRIMP models. In the IMP model some minor overlap between measurements on different skin sites occurred giving a correct classification of 92% (data not shown).

When analysing data from all individuals an overlap between the scores from different skin sites remained in all PLS-DA models (Fig. 3a–c). The inter-individual classification was improved by combining data from the NIR and IMP methods to the NIRIMP model. Even better results were obtained when the mean values of the replicates were used in the PLS-DA model (Fig. 4).

4.3 Predictability with PLS-DA

In a future diagnostic tool it is not sufficient with a good classification. It is of highest importance to predict the features correctly in an unknown measured sample. A good classification is of course a prerequisite for an accurate prediction but it might not be sufficient. Therefore, PLS-DA

was used to investigate the prediction ability in the NIR and NIRIMP models.

The predictability was basically unchanged in the two models, i.e., 72–73% correct prediction. Combining the data increased the predictability only 0.5% irrespective of the improved classification. The most plausible explanation for this is the small size of the data set which is made even smaller when one eighth of it is used as a test set.

Nevertheless, 72% correct prediction must be considered a good result since the measurements were performed on healthy skin from a homogeneous group of individuals. No big differences were, therefore, expected in skin characteristics at the various measurement sites. The majority of mispredicted values were from skin site D (the calf) that was a group that overlaps into almost every other group of skin sites. This indicated that skin site D have many physiological characteristics in common with the four other skin sites, whereas the characteristics of the other skin sites, A–C, E are less shared.

A gradient of skin thickness could be recognized in the NIRIMPred PLS-DA scatter plot (Fig. 5). The gradient goes from the thin skins at the back of hand (skin site A), the inner upper arm (skin site B), and the cheek (skin site E), through the middle thick skin on the calf (skin site D) to the thick skin on the back (skin site C). Any of the four skin sites A–C, E can of course be identified as one side of a gradient where the others are forming the other end and D always as the middle. Then, three other so far unrecognized skin characteristics might form gradients in these other directions. These results indicate that we can certainly

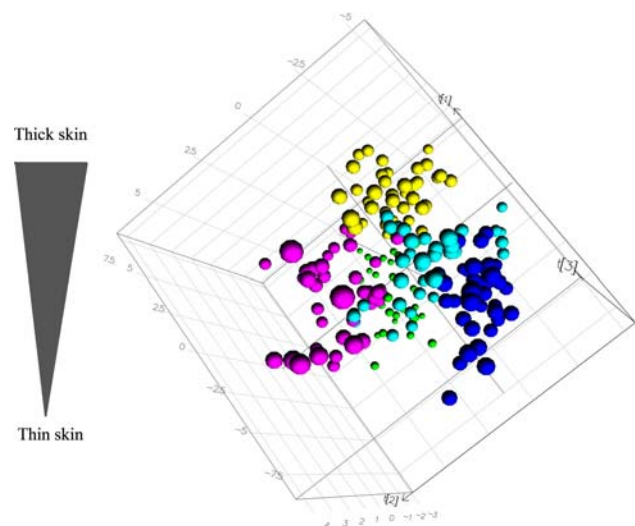


Fig. 5 The gradient of skin thickness in a NIRIMPred PLS-DA score plot with the thickest skin (site C) on the top and the thin skins (sites A, B and E) in the bottom and the middle thick skin (site D) in the centre. Skin site A: back of hand (green), B: inner upper arm (dark blue), C: back (yellow), D: calf (light blue), E: cheek (purple)

track subtle changes in skin characteristics and maybe also identify them in detail.

For the time being, if measurements from a sixth skin site with a different characteristic than the present five were added to this data set we cannot predict the corresponding scores that would be found in the three-dimensional PLS-DA score plot. The reason is that it is not known which physiological information is obtained from NIR spectroscopy and skin IMP measurements. In a previous paper [19], diabetic patients and controls were measured with Bio-IMP and NIR spectroscopy and the data was analysed with PCA. It is shown that scores from patients with prolonged neuropathy cluster in a specific direction whereas the control group and “healthy” diabetics in another. Similar results are shown in a study concerning malignant melanoma [2]. In the prediction of various illnesses, the proposed gradients of skin characteristics will, therefore, be critical. Future investigations will, for instance, focus on finding out in what direction in the multivariate space we have a gradient for increased interstitial oedema. This might benefit the understanding of the fundamentals for early diagnosis.

5 Conclusions

Healthy skin was measured with two non-invasive techniques, NIR spectroscopy and skin IMP. The data were analysed with multivariate methods and the following main conclusions were drawn.

In the practical planning of future experiments the following observations from PCA are of high importance: NIR spectroscopy measurements on skin soaked with saline solution gave basically the same results as measurements on dry skin. Skin IMP measurements are affected to a large extent by skin lotion. It was also confirmed that the two methods are robust and reliable since no block effect was observed from a long interval between two sets of measurements.

The reproducibility of the two measurement methods was analysed carefully by calculation of scatter values, i.e., SVs. It was concluded that the precision was roughly the same for the two methods with a somewhat higher precision in the NIR measurement. However, measurements on different skin sites showed a slight variation in reproducibility whereas no significant difference was observed between individuals regarding reproducibility. The later finding could be explained by the very homogeneous data.

The data were used for classification of the five different skin sites from three-dimensional PLS-DA score plots with a manual planar discriminant analysis. In an intra-individual approach the measurements from NIR provided a 100% correct classification whereas the IMP data gave

92%. An inter-individual comparison showed a large overlap between groups of measurements on different skin sites when analysing the IMP data whereas the NIR data gave 88% correct classification. When data from the two measurement methods were weighed together into one single data matrix an increase in correct classification to 93% was observed. This clearly shows the added value of combining two data sets with different measurement method. An attempt of skin site prediction with PLS-DA gave 72% correct prediction.

6 Future aspects

The results will lay the foundation in future developments of non-invasive diagnostics of various diseases. In this work the robustness of method relies on many things. Among them the most important are reproducibility in measurement and the choice of a reliable reference measurement site. In this investigation we have certainly demonstrated high reproducibility in both measurements. It is indicated that the individual differences influence the IMP data to a large extent, which could be diminished with the proposed method in this paper. In a more heterogeneous data set individual differences will probably also influence the NIR measurements.

The choice of reference skin will be of critical importance. The patient can serve as its own reference in some cases, such as diagnosis of malignant melanoma. Our primary suggestion is to use the analogous site on the opposite side of the body as a reference. However, there are some indications on a tendency to left–right side separation of the NIR data. The reason for this is at present unknown, but for development of a diagnostic method for clinical use this must be fully elucidated. In other cases a large reference database with data from both sexes and a variation of ages and anatomic body sites might be necessary. One example of this is in diagnosis of neuropathy among diabetic patients where the entire skin might be affected by the disease.

Acknowledgments The European Union Structure Foundation Objective 1 and the Unizon project NIRCE sponsored by Interreg are recognized for their financial support. Prof Paul Geladi at SLU (Unit of Biomass Technology and Chemistry, SLU Rönneby, Umeå, Sweden) is acknowledged for valuable discussions.

References

1. Aberg P, Geladi P, Nicander I et al (2005) Non-invasive and microinvasive electrical impedance spectra of skin cancer—a comparison between two techniques. *Skin Res Technol* 11:281–286
2. Aberg P, Nicander I, Hansson J et al (2004) Skin cancer identification using multifrequency electrical impedance—a potential screening tool. *IEEE Trans Biomed Eng* 51:2097–2102

3. Bodén I, Norén L, Wisten Å et al (2005) Development and optimisation of a novel skin impedance instrument. Paper presented at the 13th nordic baltic conference biomedical engineering and medical physics, Umeå, 13–17 June
4. Cobb J, Claremont D (2002) Noninvasive measurement techniques for monitoring of microvascular function in the diabetic foot. *Int J Low Extrem Wounds* 1:161–169
5. Eikje NS, Ozaki Y, Aizawa K et al (2005) Fiber optic near-infrared Raman spectroscopy for clinical noninvasive determination of water content in diseased skin and assessment of cutaneous edema. *J Biomed Opt* 10:14013
6. Eriksson L, Johansson E, Kettaneh-Wold N et al (2001) Multi- and megavariable data analysis principles and applications. Umetrics AB, Umeå
7. Friedman RJ, Rigel DS, Silverman MK et al (1991) Malignant melanoma in the 1990s: the continued importance of early detection and the role of physician examination and self-examination of the skin. *CA Cancer J Clin* 41:201–226
8. Geladi P, Kowalski BR (1986) Partial least-squares regression—a tutorial. *Anal Chim Acta* 185:1–17
9. Geladi P, Nystrom J, Eriksson JW et al (2000) A multivariate NIR study of skin alterations in diabetic patients as compared to control subjects. *J Near Infrared Spectrosc* 8:217–227
10. Har-Shai Y, Glickman YA, Siller G et al (2005) Electrical impedance scanning for melanoma diagnosis: a validation study. *Plast Reconstr Surg* 116:782–790
11. Har-Shai Y, Hai N, Taran A et al (2001) Sensitivity and positive predictive values of presurgical clinical diagnosis of excised benign and malignant skin tumors: a prospective study of 835 lesions in 778 patients. *Plast Reconstr Surg* 108:1982–1989
12. Heinemann L, Schmelzeisen-Redeker G (1998) Non-invasive continuous glucose monitoring in Type I diabetic patients with optical glucose sensors. Non-Invasive Task Force (NITF) *Diabetologia* 41:848–854
13. Kittler H, Pehamberger H, Wolff K et al (2002) Diagnostic accuracy of dermoscopy. *Lancet Oncol* 3:159–165
14. Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 47:583–621
15. Madison KC (2003) Barrier function of the skin: “la raison d’être” of the epidermis. *J Invest Dermatol* 121:231–241
16. Mahalanobis P (1936) On the generalized distance in statistics. Paper presented at the Proceedings of National Institute of Science, Calcutta
17. Martin KA (1993) Direct measurement of moisture in skin by Nir spectroscopy. *J Soc Cosmet Chem* 44:249–261
18. Nystrom J, Geladi P, Lindholm-Sethson B et al (2004) Objective measurements of radiotherapy-induced erythema. *Skin Res Technol* 10:242–250
19. Nystrom J, Lindholm-Sethson B, Stenberg L et al (2003) Combined near-infrared spectroscopy and multifrequency bio-impedance investigation of skin alterations in diabetes patients based on multivariate analyses. *Med Biol Eng Comput* 41:324–329
20. Nyström J, Svensk A-C, Lindholm-Sethson B et al (2007) Three instrumental methods for the objective evaluation of radiotherapy induced erythema in breast cancer patients and a study of the effect of skin lotions. *Acta Oncol* (in press)
21. Sjöström M, Wold S, Söderström B (1986) PLS discriminant plots in pattern recognition in practice II. Paper presented at the proceedings of PARC in practice, Amsterdam, 19–21 June 1985
22. Stahle L, Wold S (1987) Partial least squares analysis with cross-validation for the two-class problem: a Monte Carlo study. *J Chemom* 1:185–196
23. Troy TL, Thennadil SN (2001) Optical properties of human skin in the near infrared wavelength range of 1000 to 2200 nm. *J Biomed Opt* 6:167–176
24. Wiechers JW, Snieder M, Dekker N et al (2003) Factors influencing skin moisturization signal using near-infrared spectroscopy. *IFSCC Mag* 6:1–8
25. Wiernsperger NF, Bouskela E (2003) Microcirculation in insulin resistance and diabetes: more than just a complication. *Diabetes Metab* 29:6S77–87
26. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemom Intell Lab Sys* 2:37–52
27. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Sys* 58:109–130
28. Wright CI, Kroner CI, Draijer R (2006) Non-invasive methods and stimuli for evaluating the skin’s microcirculation. *J Pharmacol Toxicol Methods* 54:1–25