Med Biol Eng Comput (2012) 50:79–89
DOI 10.1007/s11517-011-0790-7

ORIGINAL ARTICLE

Outlier detection in high-density surface electromyographic signals

Hamid R. Marateb · Monica Rojas-Martínez · Marjan Mansourian · Roberto Merletti · Miguel A. Mañanas Villanueva

Received: 9 October 2010 / Accepted: 13 June 2011 / Published online: 23 June 2011

Abstract Recently developed techniques allow the analysis of surface EMG in multiple locations over the skin surface (high-density surface electromyography, HDsEMG). The detected signal includes information from a greater proportion of the muscle of interest than conventional clinical EMG. However, recording with many electrodes simultaneously often implies bad-contacts, which introduce large power-line interference in the corresponding channels, and short-circuits that cause near-zero single differential signals when using gel. Such signals are called 'outliers' in data mining. In this work, outlier detection (focusing on bad contacts) is discussed for monopolar HDsEMG signals and a new method is proposed to identify 'bad' channels. The overall performance of this method was tested using the agreement rate against three experts' opinions. Three other outlier detection methods were used for comparison. The training and test sets for such methods were selected from HDsEMG signals recorded in Triceps and Biceps Brachii in the upper arm and Brachioradialis, Anconeus, and Pronator Teres in the forearm. The sensitivity and specificity of this algorithm were, respectively, 96.9 ± 6.2 and 96.4 ± 2.5 in percent in the test set (signals registered with twenty 2D electrode arrays corresponding to a total of 2322 channels), showing that this method is promising.

Keywords Detection theory · Feature extraction · Logistic regression · Multichannel surface electromyography · Multivariate outlier detection · Robust statistics

Abbreviations

CC	Correlation coefficient
CPV	Cumulative percentage variance
EMG	Electromyography
EP	Error probability
HDsEMG	High-density surface electromyographic signals
KDE	Kernel density estimator
kNN	k-Nearest neighbors
LDOF	Local distance-based outlier factor
LOF	Local outlier factor
MAD	Median absolute deviation
MCD	Minimum covariance determinant estimator
MSD	Mahalanobis squared distance
MVIC	Maximum voluntary isometric contraction
OCA	Overall classification accuracy
PC	Principal component
PCA	Principal component analysis
PDE	Partial differential equation
PLOF	Probabilistic local outlier factor
RMS	Root mean square
SD (sd)	Standard deviation

H. R. Marateb (✉) · R. Merletti
Laboratory for Engineering of the Neuromuscular Systems, Department of Electronics, Politecnico di Torino, Turin, Italy
e-mail: hamid.marateb@polito.it

M. Rojas-Martínez
Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine, CIBER-BBN, CREB, Department ESAII, Technical University of Catalonia, UPC, Barcelona, Spain

M. Mansourian
Department of Biostatistics and Epidemiology, Health School, Isfahan University of Medical Science, Isfahan, Iran

M. A. Mañanas Villanueva
Biomedical Engineering Research Centre, CREB, CIBER-BBN, Department ESAII, Technical University of Catalonia, UPC, Barcelona, Spain

Se	Sensitivity
Sp	Specificity

1 Introduction

Some data derived from measurements may be inconsistent with others and can be viewed as outliers. Outliers affect statistical estimators. They skew the location and scale estimators (e.g., mean and covariance matrix) toward them. In the case of multiple outliers, the analysis might suggest that one or more outliers are in fact good cases (false negatives) and one or more good cases are outliers (false positives) [4].

Outlier detection is a primary step in many data mining applications associated with data quality assurance. Visual inspection of scatter plots is the most common approach to outlier detection [44]. There is a large literature on the detection of outliers mostly on the univariate case. Barnett and Lewis [3], Davies and Gather [14] and Hawkins [28] provide extensive reviews.

Outlier detection approaches can also be classified as distribution-based (Z-score and Grubb's test [24]), depth-based [9], clustering-based [15], distance-based [25, 37, 47, 50, 51], or density-based such as LOF [7], PLOF [33], and LDOF [61]. Robust statistics was also used to identify univariate outliers such as the Boxplot approach [56] and its variations [42]. The bivariate versions were also introduced in the literature as Bagplot, a Bivariate generalization of the univariate Boxplot [52], Relplot (Robust Elliptic Plot) and Quelplot (Quarter Elliptic Plot) [21]. Fuzzy expert-based methods were also used to detect outliers by combining different outlier detection methods [9].

Recording HDsEMG signals implies using several channels. During recording with many channels, it is likely to observe some low-quality signals due to poor skin–electrode contact, small electrode displacements during signal recording (movement artefacts), power-line interference, especially in monopolar recording, variations of electrode–skin impedance over time (e.g., due to intermittent or loose contacts) and loose connectors (Fig. 1 as an example). Examining the electrode–skin impedance prior to signal recording is not always practical and this impedance changes even in 1-s intervals [20]. In addition to examining the quality of the signal during recording, it is very important to identify "bad" channels, prior to off-line signal processing. Manual identification of outliers is time-consuming and depends on the expertise of the operators [34]. Thus, there is a need to design and implement automatic outlier detection systems for HDsEMG recordings.

Two automatic methods were proposed recently: [23, 39]. The first approach is a bivariate extension of Boxplot, using two-dimensional features defined by the SD of the signal in short and long epochs for each channel. The second approach is based on an expert-based Fuzzy system, and requires tuning membership functions on a training set using particle swarm optimization. It is necessary to set some thresholds empirically or based on tuning on training sets in both methods.

The objectives of this work are: (a) to extract HDsEMG features according to experts' knowledge to differentiate between "good" and "bad" channels (focusing on bad-contacts), (b) to validate feature extraction and reduction procedures statistically, and (c) to test a novel data-dependent method to estimate the cut-off threshold of outlierness factor against three other methods on the training and test sets. Detecting outliers is important for obtaining reliable EMG amplitude maps by substituting them using 2-D interpolation techniques. The proposed procedure is fully automatic, does not require any tuning step or human expert's interpretation and can also identify localized muscular activity. Preliminary results were presented recently [40].

2 Methods

2.1 Training and test set databases

Five muscles were included in the experiment: Biceps and Triceps in the upper arm and Anconeus, Brachioradialis, and Pronator Teres in the forearm. The experimental protocol consisted of isometric flexion, extension, supination and pronation at 10, 30, and 50% of the Maximal Voluntary Contraction held for 10 s. These contractions were controlled by means of a mechanical brace designed to measure isometric torques in the four directions of movement. The forearm was restrained with straps applied at the wrist. Subjects were previously trained to maintain the hand and fingers at rest during signal recording. Twelve healthy male volunteers (age, 28.3 ± 5.5 years; height: 177.8 ± 6.0 cm; weight: 75.7 ± 8.7 kg) participated in the experiment. Subjects included in the study did not have any history of neuromuscular disorders or pain or regular training of the upper limb. All subjects gave informed consent to the experimental procedure. The area of placement of the electrodes was shaved and cleaned with abrasive paste. Subjects sat in front of the mechanical brace with the back straight, the elbow joint flexed at 45°, shoulder abducted at 90° (arm parallel to sagittal plane), and forearm rotated 90° (midway between supination and pronation).

Monopolar HDsEMG signals were recorded using three 2D electrode arrays with contacts equally spaced by 10 mm in rows (y in the proximal–distal direction) and

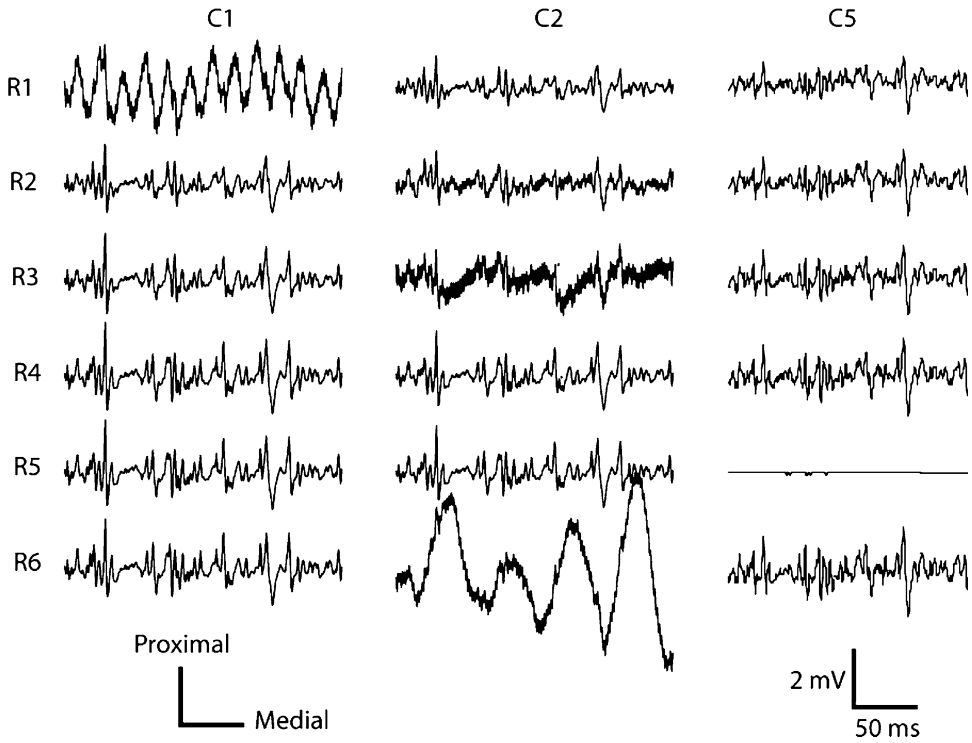


Fig. 1 Selected 250-ms monopolar HDsEMG signal epochs distributed in six rows (R1,…, R6) and three columns (C1, C2, C5) including "bad" channels from subject 1, Brachial Biceps Matrix (IED of 10 mm in both directions), elbow flexion at 10% MVIC (part of the training set no. 3). According to experts' opinion, R1C1 (CH1), R3C2 (CH9), R6C2 (CH12), and R5C5 (CH29) are "bad" channels. Although, it is possible to increase the SNR of some of them (e.g., CH9) by filtering, other channels require reconstruction using 2-D interpolation using their nearest neighbors

columns (x in the medial–lateral direction) and made of silver plated eyelets of 5 mm in diameter with the reference electrode connected to the shoulder of the subject's dominant side. Array 1 (forearm) was located 2 cm below the elbow crease with columns of electrodes (at least 4) covering the anconeus, pronator teres, and brachioradialis that were previously drawn over the surface of the skin [32]. Arrays 2 and 3 (Biceps and Triceps) consisted of 8 × 15 electrodes while Array 1 had 6 rows of electrodes and a variable number of columns (between 17 and 19) depending on the dimensions of the limb of the subject for a total of ~350 channels. Signals were recorded simultaneously by three amplifiers (EMG-USB-128 channels, sampling frequency of 2048 Hz, programmable gains of 100, 200, 500, 1000, 2000, 5000, and 10000, third-order active high-pass Butterworth filter (−3 dB cut-off frequency of 3 Hz; −18 dB/octave slope) and eighth-order switched-capacitor Bessel filter (−3 dB cut-off frequency of 710 Hz; −48 dB/octave slope); LISiN-OT Bioelettronica with synchronized sampling). Power-line interference was reduced by using a Driven Right Leg (DRL) circuit (with the DRL IN and DRL OUT connected to clavicle and wrist of the subject's dominant side, respectively). Contraction order was randomized and a rest period of 2 min was imposed between consecutive contractions. A signal "set" was obtained from each array during each contraction.

Two databases were randomly selected from the 432 recorded signal sets (recorded from 12 subjects, in four contraction types (flexion, extension, supination, and pronation), at three force levels (10, 30, and 50%), and three locations (biceps, triceps or forearm)). One database had 19 signal sets (*training)* and the other one consisted of 20 signal sets (*test*).

The training set was used to tune parameters (e.g., cut-off thresholds) of three other methods used for comparison (i.e., M1–M3). Note that the proposed method (M0) does not require tuning and its only fixed parameter, the number of nearest neighbors, was estimated theoretically (see Sect. 4.1 for details).

The number of bad channels identified by three experts (see Sect. 2.4.1 for information on combining experts opinions) was 4.05 ± 2.74 [0, 13] and 5.40 ± 3.98 [0, 16] in the training and test sets, respectively (mean ± SD) [min, max]. The number of "bad" channels in the 20 test sets was 6, 4, 7, 16, 2, 3, 12, 3, 9, 10, 7, 6, 1, 3, 4, 5, 2, 2, 6, and 0 out of 120 channels.

The minimum number of required signal sets in the training and test set databases were calculated using G*Power version 3.1.2 [17, 18]. It was required to have at least 18 signal sets for the analysis in either training or test set to preserve the statistical power of 99% at the significance level of 0.01 with the effect size of 1.3 (large effect size [10] based on the assumption made in Sect. 4.1).

2.2 Feature extraction

Experts use different strategies when identifying "bad" channels, one of them, is the identification of those that are not similar to "good" channels, considering that "good channels" usually present similar waveforms. As a

similarity measure, we may introduce the first feature Fa for channel i as $\mathrm{Fa}(i) = \mathrm{median}\{|\mathrm{CC}(x_i, x_j)|;\ j = 1\ldots n,\ i \neq j\}$, where n is the total number of recorded channels in the corresponding array, x_i are the temporal samples of channel i in a 250 ms epoch and $\mathrm{CC}(x_i, x_j)$ is the cross-correlation coefficient between waveforms x_i and x_j.

The second bivariate spectral feature, $SF = [Fb, Fc]$, is defined for each channel i in the frequency domain. First, the temporal samples of channel i in a 1 s epoch are used to calculate the power spectrum with 1 Hz resolution, P_k, $k = 1, 2, \ldots, 500$ (the DC component was *excessively* removed to satisfy one of the signal stationarity conditions [6]. High-baseline fluctuations appear in the low-frequency range ($\leq$12 Hz) while power-line interferences (with their harmonics) create isolated peaks in the spectrum. An automatic outlier detection method in the frequency domain [1] was used to identify the interferences at $P_{k>12}$. This method is based on the Hampel identifier [27] with two outputs. The first output is the total power of the interference peaks (p_o) and the second is the reconstructed power spectrum ($\tilde{p}_k$) of the signal after removing identified interferences. Now, the components of SF are defined as: $\mathrm{Fb} = (p_l/p_t)$, and $\mathrm{Fc} = (p_o/p_t)$ where $p_l = \sum_{k=1}^{12} p_k$, $p_t = \sum_{k=1}^{500} \tilde{p}_k$. Although 1 s epochs satisfy the stationarity condition required to calculate the power spectral density [30], spectral leakage [43] may enable us to use shorter epochs, e.g., 500 ms. Since the epoch length is not the same for Fa and SF features, the same SF feature is used for four consecutive values of the Fa features. The outlier detection is performed in each 250-ms non-overlapping epochs and if any of the four epochs is marked as an outlier, the corresponding recorded channel is regarded as a "bad" channel for that 1 s epoch.

The pair-wise correlations between features Fa, Fb, and Fc in a 1 s epoch of the training data sets were analyzed to check the possibility of feature reduction. Since none of the paired features in the training datasets passed one-sample Kolmogorov–Smirnov normality test [11], non-parametric Spearman's rho [13] was used to calculate the correlation between features. Significantly correlated features, 2-tailed significant level of 0.05 (95% confidence intervals) and $|\mathrm{CC}| > 0.4$ [49] in the training data set are shown in Table 1. Multiple logistic regression [35] was used to analyze the possibility of reducing the number of features. First, Z-scores [45] were used to identify possible outliers. Z-scores with an absolute value greater than 2.5 were labeled as potential outliers [3]. The result of this outlier detection approach was compared with that of experts' opinions resulting in a binary value of 1 for agreement and 0 for disagreement. Multiple logistic regressions enabled us to identify dominant features (2-tailed significant level of 0.05) in the model as listed in Table 1. Dominant features are distributed in different sets indicating that no feature can be generally omitted from the analysis. The number of dominant features in some of the training data sets is more than one, indicating that univariate outlier detection methods are not suitable for our application. Robust PCA based on Projection Pursuit [12, 55, 58] was also used to identify PC's of Fa, Fb, and Fc features covering the CPV [38] of 95%. The number of PC's is also shown in Table 1 as the minimum number of features to use. 2-D representation of the first two uncorrelated features in training set no. 3 is shown in Fig. 2 indicating the compact representation of "bad" channels using the recommended number of PC's (to cover CPV of 95%) in Table 1. Since the training set was chosen randomly and was verified by the Runs test at the 5% significance level [29], it is possible to generalize the following results to the whole data set:

1. All of the proposed features, i.e., Fa, Fb, Fc must be considered. After uncorrelating them using Projection Pursuit, which is robust to outliers, those transformed features covering the CPV of 95% are used.
2. Univariate Outlier detection methods are not appropriate for our application since the number of dominant features is more than one in most cases.

2.3 Outlier detection

LDOF [61] was used to identify outliers. LDOF requires the number of nearest samples as input. We used 24 neighbors for this classifier (refer to Sect. 4.1 for the discussion about selecting the number of neighbors). This method reports the degree of outlierness of an object instead of a binary decision, requiring a cutoff value. A data-driven threshold based on adaptive kernel density estimation [59] was used as the cutoff threshold.

2.3.1 Classifier

Multidimensional uncorrelated features (PC's covering CPV of 95%) were used as an input to the LDOF method. Suppose that k, the number of nearest samples, is 24 (see Sect. 4.1), g is the feature and N_p is a set including kNN of the measured feature g_p, then LDOF for each object g_p is calculated using Eq. 1:

$$\bar{d}_{g_p} := \frac{1}{k}\sum_{g_i \in N_p} \mathrm{dist}(g_i, g_p),\ \bar{D}_{g_p} = \frac{1}{k(k-1)}\sum_{g_i, g_{i'} \in N_p, i \neq i'} \mathrm{dist}(g_i, g_{i'}),\ \mathrm{LDOF}_k(g_p) = \frac{\bar{d}_{g_p}}{\bar{D}_{g_p}} \quad (1)$$

where $\mathrm{dist}(g_i, g_{i'}) \geq 0$ is a distance measure between objects g_i and $g_{i'}$, $\bar{d}_{g_p}$ is kNN distance of g_p, i.e., the average of the distances from g_p to all objects in N_p and

Table 1 The correlation analysis of the features in the training data set. AB, BC, and AC stand for significant correlation (2-tailed significant level of 0.05 and $|CC| > 0.4$) between (Fa, Fb), (Fb, Fc), and (Fa, Fc)

Output	Training set																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Significantly correlated features	AB AC	–	BC	BC	AB BC AC	BC	AC	AC BC	AB	–	–	–	AB	–	AB	AB BC AC	–	–	AB
Dominant features	Fa	Fa Fb	Fa	Fa Fc	NA	–	–	–	Fa	Fb	NA	Fa	Fc	–	–	Fb Fc	–	Fa Fb	Fa Fc
Number of PC's	1	3	2	2	1	2	3	2	3	2	2	2	3	2	2	3	3	3	3

Dominant features were identified using Logistic regression (2-tailed significant level of 0.05). NA stands for not available, where there was 100% agreement between the multivariate Grubb's test and that of expert's opinion, resulting in no regression error to analyze. Number of principal components (PC's) covering the cumulative percentage variance (CPV) [38] of 95% is also shown as the minimum number of features to use

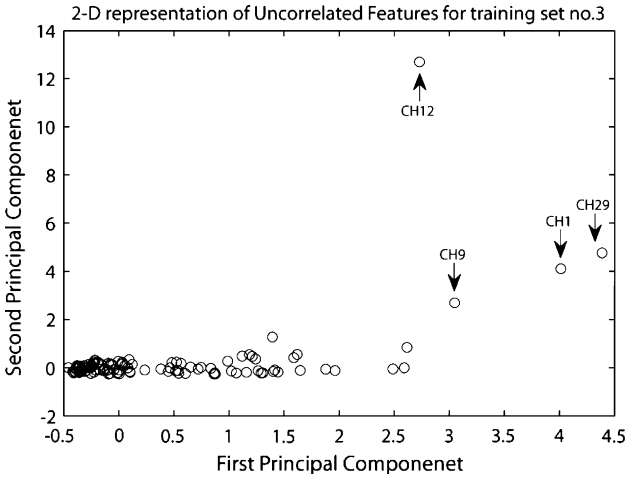


Fig. 2 2-D representation of the first two PC's in training set no. 3, elbow flexion at 10% MVIC, whose corresponding HDsEMG signals are shown in Fig. 1. Four outliers identified by the experts (recording channels no. 1, 9, 12, and 29) are marked with *arrows*. According to the number of PC's proposed in Table 1, these two first components are compact features

$\bar{D}_{g_p}$ is kNN inner distance of g_p, i.e., the average of the distances among objects in N_p. LDOF captures the degree to which object g_p deviates from its neighbors. The squared Euclidean distance $\left(\|\|^2\right)$ was used for distance measure.

The output of this classifier is the degree of outlier-ness of objects in scattered datasets. The output of this classifier for the training sets no. 3 and 9 is shown in Fig. 3a and b, respectively.

2.3.2 Accurate classification boundary

Calculating the distribution of LDOF values, it is possible to isolate the bulk of the data. Since the distribution of the data is not known, it is necessary to use non-parametric density estimation methods. Among them, KDE was used since it does not present drawbacks existing in histograms, such as high sensitivity to the number of bins and discontinuities [59].

If $x_1, x_2, \ldots, x_n$ are i.i.d. (independent and identically distributed) samples of a random variable with probability density function f, such as LDOF values, then the kernel density approximation of its probability density function (KDE) is calculated using Eq. 2.

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{2}$$

where K is a kernel function (e.g., normal, triangular), h is a smoothing parameter (bandwidth), and n is the number of samples. The performance of KDE, in representing the probability density function of x, is dependent on the choice of the kernel function and bandwidth. In our case, the Epanechnikov kernel function was chosen defined in Eq. 3.

$$K(u) = \frac{3}{4}\left(1 - u^2\right)\mathrm{I}_{\{|u| \le 1\}}(u) \tag{3}$$

where $I_A(u)$ is the indicator function defined as 1 if $u \in A$ and 0 elsewhere. This Kernel function was used since it has the highest efficiency (=1) among others because it minimizes asymptotic mean integrated squared error (AMISE) which is one of the performance measures of KDE [54].

Instead of using a fixed bandwidth, as in the first generation methods mentioned in [31], an adaptive data-driven bandwidth was chosen because it is necessary to adapt the bandwidth to the local density. An adaptive KDE adapts to the sparseness of the data by using a broader kernel over observations located in regions of low density (varying the bandwidth inversely with the density) [57]. Following the "Solve-the-Equation Plug-In-Approach" (second generation methods [31]) to calculate the adaptive KDE, it is first required to estimate a pilot density ($\tilde{f}$) and bandwidth ($\tilde{h}$).

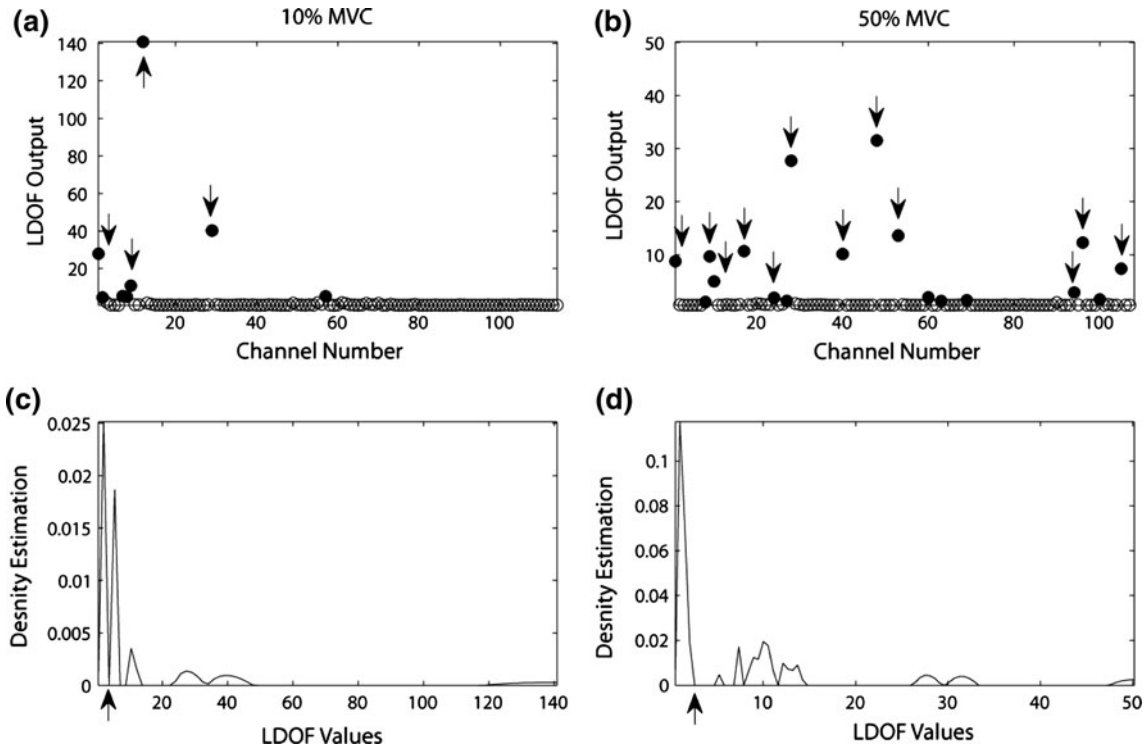


Fig. 3 The output of LDOF classifier for each channel using PC's in training set no. 3 whose signals and features were shown in Figs. 1 and 2 respectively, for elbow flexion at 10% MVIC (**a**), and training set no. 9 recorded from subject 3, Forearm Matrix (IED of 10 mm in both directions), for elbow flexion at 50% MVIC (**b**). Outliers identified by the experts are marked with *arrows* while those identified by the automatic outlier detection method are marked with *filled circle*. Kernel density estimations of LDOF values (with adaptive bandwidth and Epanechnikov Kernel function, see Sect. 2.3.2) for training sets no. 3 and 9 are shown in **c** and **d**, respectively. The borderlines of the bulk of the LDOF values are marked with *arrows* with the LDOF values of 3.8 and 2.58, respectively

Using Epanechnikov kernel function and following "The Rules of Thumb" [54], it is possible to estimate $\tilde{f}$ and $\tilde{h}$. This density estimation is oversmoothed and therefore suitable for the next tuning step based on the local density. The adaptive KDE is given by $\hat{f}_{\text{adp}}(x)$ and defined as follows [57]:

$$\hat{f}_{\text{adp}}(x) = \sum_{i=1}^{n} \frac{1}{n\tilde{h}\lambda_i} \mathrm{K}\left(\frac{x - x_i}{\tilde{h}\lambda_i}\right) \tag{4}$$

where x_i's are the data points (LDOF values), K is the Epanechnikov kernel function, n is the number of data points, and λ_i's (local bandwidth factors) are defined as $\lambda_i = \lambda(x_i) = \sqrt{G/\tilde{f}(x_i)}$. G is the geometric mean over all i of the pilot density estimate $\tilde{f}(x_i)$. In this approach, h controls the overall degree of smoothing while the λ_i values stretch or shrink the sample point's bandwidth to adapt to the density of the data (see Fig. 3c, d).

The first local minimum of the KDE was used as the margin of the bulk of the data whose LDOF value was used as the cutoff point for the classifier. Since kernel smoothing is used, the first valley of $\hat{f}_{\text{adp}}\left(\text{LDOF}_k\left(g_p\right)\right)$ can be identified as the first local minimum instead of using other complicated automatic valley-detection approaches [36]. The proposed threshold detection method is indeed the 1-d interpretation of the histogram-based image segmentation [53] that has been extensively used in machine learning. The cut-off threshold selection procedures for the third and ninth training sets are shown in Fig. 3.

2.4 Validation of the outlier detection method

2.4.1 Gold standard

Outliers were manually detected by three experts in the training and test sets containing 19 and 20 signal sets (at least 108 channels in each set). One second of the HDsEMG signals in the plateau force region in each dataset was analyzed. The "MODE" operator that is the majority vote expressed as binary value 0 for good channels and 1 for outliers was used to combine three experts' opinions to identify artifacts for each channel of each set as the gold standard. Reliability of agreement between experts was assessed using Fleiss' Kappa index [19] and scored 88.96 and 83.92% pointing to "almost perfect agreement" for the training and test set, respectively.

2.4.2 Performance indices

Three different performance measures were used to assess the matching of the proposed outlier detection method in each set with the "gold" standard. The first index was OCA, the overall classification accuracy [16] as the percentage of agreement between the result of classifier and that of expert's opinion. The second and third indices were the sensitivity (Se) and specificity (Sp) of the outlier detection algorithm. Considering true positive (TP) and true negative (TN) as the number of correctly identified outliers and "good" channels respectively, false positive (FP) as the number of "good" channels identified as outliers and false negative (FN) as the number of outliers identified as "good" channels, it is possible to define our performance indices as:

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \mathrm{OCA} = \mathrm{Acc} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FN}+\mathrm{FP}},$$

and

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$$

where Se, Sp, and Acc are sensitivity, specificity, and accuracy as the capability of the outlier detection algorithm to correctly identify outliers, preserve "good" channels and the overall performance of the classifier. They provide a compact representation of the performance of the method.

2.4.3 Comparison with other outlier detection methods

The performance of the proposed outlier detection method (M0) was compared with that of three methods (M1–M3) selected because of their superior performance in comparison with other approaches proposed in the review literature (e.g., [26]). These methods are listed below:

M1 PC's were used to calculate probabilistic local outlier factor (PLOF) [33] (number of nearest neighbors $k = 24$) and the cut-off was estimated using Hampel method [1, 60] with the coefficient of 5.2 [46]. PLOF approach was selected because of its high performance with respect to other density-based outlier detection methods, e.g., LOF [7], kNN, and weighted kNN [2].

M2 PC's were used to calculate Robust Mahalanobis distance using fast MCD method [51]. The robust distances d_i^2 (for each channel i) were transformed to new distances D_i^2 according to the following equation, $D_i^2 = \chi^2_{p,0.5}\frac{d_i^2}{\mathrm{median}(d_i^2)}$ where p is the degrees of freedom, i.e., the number of PC's. This transformation is aimed at matching the midpoints (medians) of the theoretical χ^2 distribution with the empirical distances [41]. Then, the cut-off point of $\sqrt{\chi^2_{p,0.975}}$ was used for D_i distances [50].

M3 Z-scores were used to identify possible outliers using [Fa, Fb, Fc] features and cut-off point of 2.5 [3]. This method was the only univariate outlier detection method used in this study.

3 Results

3.1 Performance against the gold standard

Overall performance indices of the proposed outlier detection method (M0) and of three other implemented approaches (M1–M3) are listed in Table 2 for the training and test sets (suffix r and t are related with the training and test sets, respectively, in the table). Se and Sp of methods M0, M1, M2, and M3 are shown in Fig. 4 for each test set.

Figure 4 shows that M0 and M2 have much higher sensitivity (Se) in comparison with M1 and M3. Although Se index of M2 is slightly better than that of M0 (<3% in average), its specificity (Sp) index is always much less than M0 (>13% in average). Using Wilcoxon rank test [48] Se was not significantly different between M0 and M2, but Sp of M0 was significantly superior to M2 (at the significance level of 0.05) indicating that M2 results in more false positives (i.e., labels more good channels as outliers) in comparison with M0. Since the number of "good" channels is usually much higher than "bad" channels, and avoiding false negatives has more importance than reconstructing few more false positives not to misinterpret the data, we might conclude that M2 and, especially, M0 are preferred among the methods studied.

3.2 EMG activity map

After identifying outliers in a 2-D array, it is possible to reconstruct the average RMS activity map using 2-D interpolation methods to replace the "bad" channel. This procedure is shown in Fig. 5 (test set no. 7) before and after identifying the outliers and correcting the image by interpolation using partial differential equation (PDE)-based image inpainting method [5].

4 Discussion

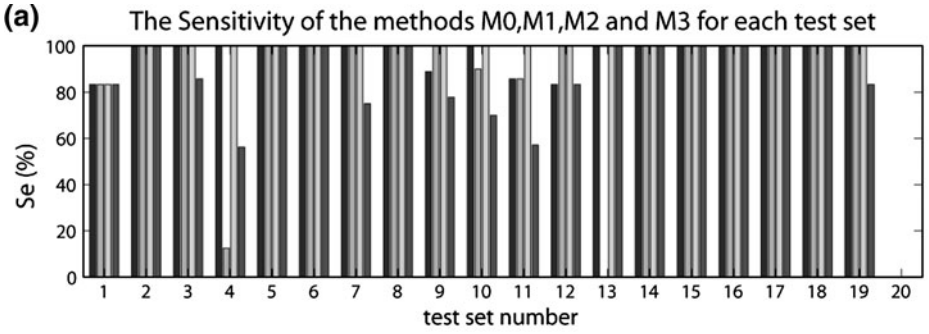
4.1 Assumptions

We assume that the percentage of outlier channels ($P_{\max}$) is not more than 25%, e.g., max. 27 "bad" channels in 108

Table 2 Overall performance indices (sensitivity (Se), specificity (Sp), and overall classification accuracy (OCA)) of the proposed outlier detection method (M0) compared with three other implemented approaches (M1–M3) for the analyzed data sets (suffix r and t, respectively, indicate the training and test set)

Performance (%)	Outlier detection methods			
	M0	M1	M2	M3
Se_r	89.1 ± 15.7	85.2 ± 19.6	89.2 ± 15.2	81.8 ± 14.6
Sp_r	99.5 ± 0.7	95.9 ± 3.3	92.2 ± 3.9	97.2 ± 3.7
OCA_r	98.9 ± 1.1	91.2 ± 3.3	92.1 ± 3.9	93.9 ± 5.2
Se_t	96.9 ± 6.2	87.9 ± 29.3	99.1 ± 3.8	87.9 ± 15.0
	[83.3, 100]	[83.3, 100]	[83.3, 100]	[56.3, 100]
Sp_t	96.4 ± 2.5	96.8 ± 3.1	83.1 ± 6.4	99.2 ± 1.3
	[90.1, 100]	[81.8, 95.8]	[69.2, 92.8]	[95.0, 100]
OCA_t	91.9 ± 4.1	93.2 ± 4.2	72.9 ± 6.5	95.7 ± 1.5
	[82.4, 98.3]	[75.0, 95.8]	[69.2, 90.4]	[92.5, 98.3]

Performance measures (mean ± sd, [min, max]) are listed in percent



		Automatic Detection	
		Artifact	Non artifact
Experts' Knowledge	Artifact	TP	FN
	Non Artifact	FP	TN

(c)

Method	TP	TN	FP	FN
M0	5.1 ± 3.9 (4.0)	109.3 ± 8.4 (111.5)	1.4 ± 1.6 (1.0)	0.3 ± 0.6 (0.0)
M1	5.3 ± 4.0 (4.5)	102.3 ± 6.3 (102.5)	12.4 ± 4.3 (12.0)	0.1 ± 0.2 (0.0)
M2	5.3 ± 4.0 (4.5)	91.9 ± 8.8 (92.5)	18.8 ± 7.7 (17.0)	0.1 ± 0.2 (0.0)
M3	4.9 ± 3.2 (4.5)	107.3 ± 7.0 (109.0)	7.3 ± 8.9 (6.0)	0.5 ± 0.8 (0.0)

(d)

Fig. 4 The performance measures of the methods M0 (the proposed one), M1, M2, and M3 for each test set shown consecutively categorically in the *bar graph* (Se, the percentage of correctly identified outliers, and Sp, the percentage of correctly identified "good" channels, in **a** and **b**, respectively out of 20 sets). The definition of indices used in signal detection theory: true positive (TP), false negative (FN), false positive (FP), and true negative (TN) (**c**) and the value of those indices of the whole test set for the methods analyzed in mean ± sd(median) format (**d**). Since test set no. 20 did not contain any outlier, and the results of neither of the methods had FN in this set, Se was not defined for this set. Although the Se index of M2 is slightly better than M0 (<3% in average), its Sp index is significantly less than for M0 (at the significance level of 0.05)

recorded channels. In our database, the maximal observed percentage was 12%. The probability of having more than 11 outliers in the kNN set ($k = 24$, $P_{\max} = 25\%$) is 0.72% = EP. When k is set to 48, EP decreases to *0.02%* but it decreases the localized-activity identification performance of the method. Decreasing k to 8, on the other

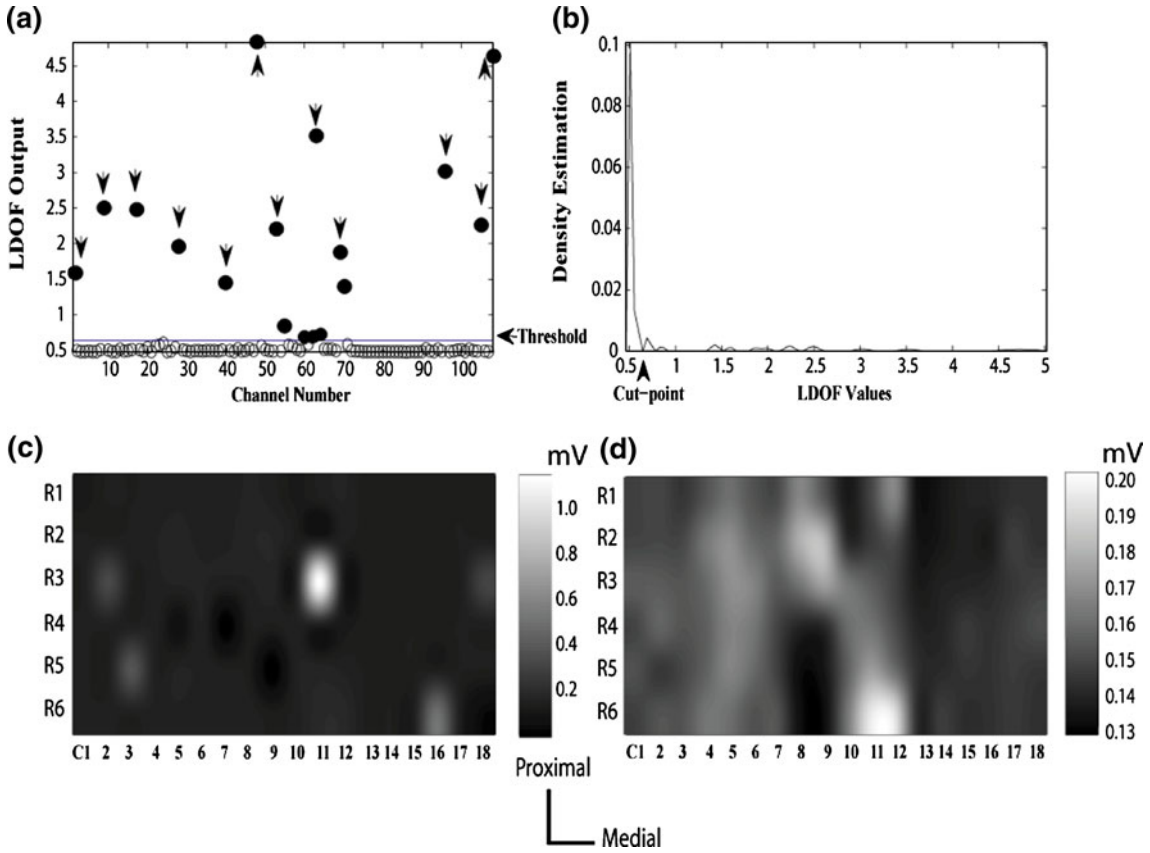


Fig. 5 The output of LDOF classifier for each channel using three PC's in test set no. 7 recorded from subject no. 13, Forearm Matrix (IED of 10 mm in both directions), elbow flexion at 10% MVIC (**a**). Outliers identified by the experts are marked with *arrows* while those identified by the automatic outlier detection method are marked with *filled circle*. KDE of LDOF values is shown in **b**. The first local minimum (0.65) was used as the threshold to detect outliers. Interpolated monopolar amplitude map (RMS) computed for 60-ms epoch before (**c**) and after (**d**) outlier detection and removal procedure. The *x*-axis and *y*-axis are, respectively, array columns and rows (IED of 10 mm in both directions). The original frame had "bad" channels (R1C1 (CH1), R3C2 (CH9), R5C3 (CH17), R4C5 (CH28), R4C7 (CH40), R6C8 (CH48), R5C9 (CH53), R3C11 (CH63), R3C12 (CH69), R6C16 (CH96), R3C18 (CH105), and R6C18 (CH108)) that were identified by our proposed outlier detection method and interpolated afterwards. The image histograms were equalized for contrast enhancement and magnified for clarity [22]. Thus, each activity map has its own *grayscale color bar* in mV. Five "good" channels were also labeled as outliers because the estimated threshold was not perfect, which are R1C10 (CH55), R6C10 (CH60), R2C11 (CH62), R4C11 (CH64), and R4C12 (CH70). They were also reconstructed using neighboring channels. The original activity of the muscles beyond the forearm matrix, were disclosed after image inpainting

hand, increases EP to 11.38% deteriorating the performance of the algorithm. EP for our database ($P_{\max} = 12\%$) was *0.0006%* which is acceptable.

4.2 Other outlier detection methods

Although our method and M2 have similar Se value, our method *always* preserves more 'good' channels than method M2 (significantly higher Sp value). In addition, M2 has the computational complexity of $O(n^3)$ (n as the number of recorded channels). The proposed method, M0, has the computational complexity of $O(n \log(n))$ which is more efficient. In practice, analyzing the whole test sets (2322 signals) took 4.6 and 35.5 s for the methods M0 and M2, respectively, on a dual core Intel® CPU processor (1.83 GHz) with 2.99 GB of RAM using Matlab® 7 program. Considering the accuracy and efficiency, our proposed method is the most promising approach among others analyzed. Although method M3, the only univariate method used in this study, had lower Se range than M0 and M2 as expected from Sect. 2.2, it can identify extreme-case outliers in each feature dimension. The method M3 has higher false negatives than M0 and cannot identify outliers similar to "good" channels. Since it is easy to implement and has the computational complexity of $O(n)$, it can be a suitable choice for pre-processing or on-line implementation as a substitution of Boxplot method.

4.3 Final considerations

Results showed that this method is reliable to identify outliers and preserve "good" channels better than the other

methods. However, there are some considerations: (1) the electrode–skin impedance was not measured. This measurement might be valuable when analyzing impedance change in the recorded signal. However, our approach can be used even if it is not possible to measure this quantity during recording, (2) the method, was designed for HDsEMG signal and the input features might not be suitable for other biopotential signals specially those whose practical bandwidth includes very low frequencies (<2 Hz), (3) in case of different electrodes (e.g., dry electrodes that have more bad-contacts) and muscles (e.g., small muscles), $P_{\max}$ has to be estimated based on the pilot study to estimate the number of kNN's in Sect. 4.1. Although other amplifiers were not used in our experiment, different high-pass filtering cannot completely attenuate movement artefacts and does not affect the proposed data-driven cut-off threshold estimation method. Moreover, our proposed method does not require tuning to depend on the particular data. However, the low-frequency feature extraction must be evaluated and possibly adapted for other amplifiers, and (4) raw (unfiltered) monopolar HDsEMG signals were analyzed. Higher order spatial filtering might have a different outcome.

Acknowledgments We are grateful to Kevin McGill for reviewing a draft of this paper. This work was supported by Compagnia di San Paolo, Fondazione CRT, the Spanish government (TEC2008-02754) and the Doctoral School of Politecnico di Torino, Italy.

References

1. Allen DP (2009) A frequency domain hampel filter for blind rejection of sinusoidal interference from electromyograms. J Neurosci Methods 177:303–310
2. Angiulli C, Pizzuti (2002) Fast outliers detection in high dimensional spaces. In: 6th European conference on principles and practice of knowledge discovery in databases (PKDD'02), August 19–23, Helsinki, Finland, pp 15–26.
3. Barnett V, Lewis T (1994) Outliers in statistical data, 3rd edn. Wiley, New York
4. Ben-Gal I (2005) Outlier detection. In: Maimon O, Rockach L (eds) Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers. Kluwer Academic Publishers, Boston ISBN 0-387-24435-2
5. Bertalmio M (2006) Strong-continuation, contrast-invariant inpainting with a third-order optimal PDE. IEEE Trans Image Process 15(7):1934–1938
6. Bilodeau M, Cincera M, Arsenault AB, Gravel D (1997) Normality and stationarity of EMG signals of elbow flexor muscles during ramp and step isometric contractions. J Electromyogr Kinesiol 7(2):87–96
7. Breunig MM, Kriegel H, Ng RT, Sander J (2000) LOF: identifying density-based local outlier. In: Proceedings of the ACM SIGMOD international conference on management of data, May 14–19, Dallas, TX, USA, pp 93–104
8. Cateni S, Colla V, Vannucci M (2008) Outlier detection methods for industrial applications. In: Juan Manuel Ramos Arregui (ed) Advances in robotics, automation and control, I-tech education and publishing, pp 265–282
9. Cateni S, Colla V, Vannucci M (2009) A fuzzy system for combining different outliers detection methods. In: Proceedings of the IASTED international conference on artificial intelligence and applications (AIA 2009), February 16–18, Innsbruck, Austria
10. Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum Associates, New Jersey
11. Conover WJ (1980) Practical nonparametric statistics, 2nd edn. Wiley, New York
12. Croux C, Ruiz-Gazen A (2005) High breakdown estimators for principal components: the projection-pursuit approach revisited. J Multivar Anal 95(1):206–226
13. Daniels HE (1950) Rank correlation and population models. J R Stat Soc B 12:171–181
14. Davies L, Gather U (1993) The identification of multiple outliers. J Am Stat Assoc 88:782–792
15. Duan L, Xu L, Liu Y, Lee J (2009) Cluster-based outlier detection. Ann Oper Res 168(1):151–168
16. Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley, New York
17. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Method 39(2):175–191
18. Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav Res Methods 41(4):1149–1160
19. Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76(5):378–382
20. Geddes LA, Valentinuzzi ME (1973) Temporal changes in electrode impedance while recording the electrocardiogram with dry electrodes. Ann Biomed Eng 1:356–367
21. Goldberg KM, Iglewicz B (1992) Bivariate extensions of boxplot. Technometrics 34(3):307–320
22. Gonzalez RC, Woods RE (2002) Digital image processing, 2nd edn. Prentice-Hall, New Jersey
23. Grönlund C, Roeleveld K, Holtermann A, Karlsson JS (2005) Online signal quality estimation of multichannel surface electromyograms. Med Biol Eng Comput 43(3):357–364
24. Grubbs F (1969) Procedures for detecting outlying observations in samples. Technometrics 11(1):1–21
25. Hadi AS (1992) Identifying outliers in multivariate data. J R Stat Soc B 54(3):761–771
26. Hadi AS, Rahmatullah Imon AHM, Werner M (2009) Detection of outliers. WIREs Comp Stat 1. Wiley, New York, pp 57–70
27. Hampel FR (1971) A general qualitative measure of robustness. Ann Stat 42:1887–1896
28. Hawkins DM (1980) Identification of outliers. Chapman and Hall, London, New York
29. Hosmer DW, Hollander M, Wolfe DA (1973) Nonparametric statistical methods, 3rd edn. Wiley, New York
30. Inbar GF, Noujaim AE (1984) On surface EMG spectral characterization and its application to diagnostic classification. IEEE Trans Biomed Eng 31(9):597–604
31. Jones MC, Marron JS, Sheather ST (1996) A brief survey of bandwidth selection for density estimation. J Am Stat Assoc 91(433):401–407
32. Kendall FP, McCreary EK, Provance PG (1993) Muscles and testing function, 3rd edn. Williams & Wilkins, Baltimore
33. Kriegel H, Kröger P, Schubert E, Zimek A (2009) LoOP: local outlier probabilities. ACM, CIKM'09, Hong Kong
34. Last M, Kandel A (1999) Automated perceptions in data mining. In: Proceedings of the 8th IEEE international conference on fuzzy systems, August 22–25, Seoul, Korea. Part I: pp 190–197. doi: 10.1109/FUZZY.1999.793233
35. Lemeshow S (2000) Applied logistic regression, 2nd edn. Wiley, New York

36. López AM, Lumbreras F, Serrat J, Villanueva JJ (1999) Evaluation of methods for ridge and valley detection. IEEE Trans Pattern Anal 21(4):327–335
37. Mahalanobis PC (1936) On the generalized distance in statistics. Proc Nat Inst Sci India 2:49–55
38. Malinowski ER (1991) Factor analysis in chemistry, 2nd edn. Wiley-Interscience, New York
39. Marateb HR, Soares AF, Rojas M, Merletti R (2009) An expert-based fuzzy system for automatically identifying the location of muscle innervation zones in surface electromyography. In: Proceedings of the International Xth Quantitative EMG conference, May 6–10, Venice, Italy
40. Marateb HR, Rojas-Martínez M, Mañanas Villanueva MA, Merletti R (2010) Robust outlier detection in high-density surface electromyographic signals. Conf Proc IEEE Eng Med Biol Soc 1:4850–4853
41. Maronna RA, Zamar RH (2002) Robust estimates of location and dispersion of high-dimensional datasets. Technometrics 44(4):307–317
42. Mcgill R, Tukey JW, Larsen WA (1978) Variations of box plots. Am Stat 32(1):12–16
43. Mewett DT, Reynolds KJ, Nazeran H (2004) Reducing power line interference in digitized electromyograms recordings by spectrum interpolation. Med Biol Eng Comput 42:524–531
44. Minium EW, Clarke RB, Coladarci T (1999) Elements of statistical reasoning. Wiley, New York
45. Mosteller F, Tukey JW (1977) Data analysis and regression: a second course in statistics, 1st edn. Addison, Wesley
46. Pearson RK (2002) Outliers in process modeling and identification. IEEE Trans Control Syst Technol 10:55–63
47. Penny KI, Jolliffe IT (2001) A comparison of multivariate outlier detection methods for clinical laboratory safety data. Statistician 50(3):295–308
48. Randles RH (1988) Wilcoxon signed rank test. In: Kotz S, Johnson NL (eds) Encyclopedia of statistical sciences, vol 9. Wiley, New York, pp 613–616
49. Rodriguez RN (1982) Correlation. In: Kotz S, Johnson NL (eds) Encyclopedia of statistical sciences, vol 2. Wiley, New York, pp 193–204
50. Rousseeuw PJ, Leory AM (1987) Robust regression and outlier detection. Wiley, New York ISBN: 9780471852339
51. Rousseeuw PJ, Van Driessen K (1999) A fast algorithm for the minimum covariance determinant estimator. Technometrics 41:212–223
52. Rousseeuw PJ, Ruts I, Tukey JW (1999) The bagplot: a bivariate boxplot. Am Stat 53(4):382–387
53. Shapiro LG, Stockman GC (2001) Computer vision. Prentice-Hall, New Jersey ISBN 0-13-030796-3
54. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
55. Small CG (1990) A survey of multidimensional medians. Int Stat Rev 58(3):263–277
56. Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading, MA
57. Van Kerm P (2003) Adaptive kernel density estimation. Stat J 3(2):148–156
58. Varid Y, Zhang CH (2000) The multivariate L1-median and associated data depth. Proc Natl Acad Sci USA 97(4):1423–1426
59. Wand MP, Jones MC (1994) Kernel smoothing (Monographs on statistics applied probability). 1st ed., Chapman Hall/CRC, London, p 71
60. Westerweel J (1994) Efficient detection of spurious vectors in particle image velocimetry data. Exp Fluids 16:236–247
61. Zhang K, Hutter M, Jin H (2009) A new local distance-based outlier detection approach for scattered real-world data. Lecturer Notes in Computer Science, Advances in knowledge discovery and data mining, 13th Pacific-Asia Conference, PAKDD 2009, April 27–30, Bangkok, Thailand, vol 5476, pp 813–822. doi: 10.1007/978-3-642-01307-2_84