Chemometrics and Intelligent Laboratory Systems 97 (2009) 132–140

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Automatic preprocessing of electrophoretic images

M. Daszykowski [a], M.S. Wróbel [a], A. Bierczynska-Krzysik [b], J. Silberring [b], G. Lubec [c], B. Walczak [a,*]

[a] Department of Chemometrics, Institute of Chemistry, Silesian University 9 Szkolna Street, 40-006 Katowice, Poland
[b] Department of Neurobiochemistry, Faculty of Chemistry, Jagiellonian University, 3 Ingardena Street, 30-060 Krakow, Poland
[c] Department of Pediatrics, Medical University of Vienna, Waehringer Guertel 18, A-1090 Vienna, Austria

ARTICLE INFO

Article history:
Received 8 October 2008
Received in revised form 3 March 2009
Accepted 5 March 2009
Available online 13 March 2009

Keywords:
Comparative proteomics
Gel electrophoresis
Significant features
Images analysis
Robust regression
Robust orthogonal regression

ABSTRACT

Analysis of two-dimensional (2D) electrophoretic images is a multi-step approach, enabling application of a variety of methods at different stages of data processing. The choice of these, as well as input parameters, leads to software-induced variations. Effective preprocessing methods, which do not require optimization of input parameters, are potent in eliminating software-induced variations. As a general method for background elimination and image scaling, robust Orthogonal Regression (rOR) is proposed and compared with Orthogonal Regression. This comparison is based on the univariate and multivariate approaches of feature selection, exploring the idea developed for significance analysis of microarray data [V. Goss Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, P. Natl. Acad. Sci. U. S. A., 98 (2001) 5116–5121] and adapted to the analysis of proteomic data. All calculations are performed at the pixel level.

1. Introduction

The main goal of comparative proteomics is to define differences in the expression level of proteins at different biological states (e.g., healthy versus diseased, or control versus treated). Although gel electrophoresis is considered as a core analytical technique for the separation and quantification of proteins, processing of 2D electrophoretic images remains the main bottleneck of the proteomic studies. As demonstrated in [2], commercially available software packages introduce high variability to the analyzed data, mainly at the stages of data pretreatment (e.g., normalization, scaling) and preprocessing (e.g., smoothing, background elimination). Another problem faced during analysis of gel images is caused by requirement of spots detection. It is very rare that single spot represents single protein. In a majority of cases, we have to deal with overlapping spots [3]. Their automated detection introduces substantial errors, thus affecting the final statistical analysis. As demonstrated in [4], this is the cause of many missing elements in the spots table.

These problems can be eliminated by introducing a more objective approach to the data pretreatment and releasing restrictions for spot detection and fitting parameters. In our start-to-end approach [5–12], we proposed working at the pixel, and not at the spot level, in order to avoid problems with missing elements in the data table. At the stage of image pretreatment, we opted for background removal using the rolling ball or the penalized asymmetric least squares approach [13,14]. These two methods are very effective, but they require optimization of input parameters. Although, for scientists familiar with principles of the aforementioned methods this optimization is relatively easy to perform, the approach may prove a serious drawback for inexperienced users of the software. To make the whole proteomic analysis more objective, it is recommended to replace the background removal techniques by regression methods, which automatically account for the background of masked images. In our study, two regression methods were tested and compared based on the false discovery rate (FDR), i.e., on the percentage of pixels identified by chance as significantly differentiating two classes of samples (control and treated). Method developed for the significance analysis of the microarray data, SAM [1], was now adapted for the current purpose. It allows adjusting of the significance level of test of multiple variables and it was found very attractive for the analysis of the proteomic data as well.

2. Theory

2.1. Background

The original electrophoretic images are noisy, with artifacts, a significant background, saturated spots and local distortions of the spots (proteins) position. To detect proteins which significantly differentiate the studied classes of samples, all these undesirable effects have to be eliminated. There are many alternative approaches to each step of image analysis. For instance, images can be warped using manually selected markers or the automatic area or feature-based approaches [15–32], they can be de-noised using different filters [8], etc. In different software packages, various methods of

* Corresponding author.
E-mail address: beata@us.edu.pl (B. Walczak).

0169-7439/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2009.03.002

warping and preprocessing are implemented. However, with the majority of software packages, spots detection is required and it is performed based on the watershed algorithm. Once spot detection is completed, the original images are replaced by so-called ideal images, with each spot represented by an ideal Gaussian fitted to the data. This approach is effective when individual spots represent individual proteins. In the case of spot overlapping, certain problems can arise with data fitting (leading to missing elements in the data table [4]). Moreover, spots identified in the area of the overlap can be significantly influenced by an applied background elimination method and/or its input parameters (see Fig. 1).

In the start-to-end approach [5], step of spot detection is avoided. Automatic warping of normalized images (fuzzy warping) requires identification of the local maxima only, and all calculations are performed at the pixel level, and not at the spot level [5,12]. For the background elimination, the penalized asymmetric least squares or the rolling ball approach was proposed [13,14]. Once the background is removed, all images are used for calculation of the master image, from which the binary mask can be constructed.

Application of the binary mask considerably reduces the number of pixels to be further processed. Using the binary mask for the individual images, we are further considering the intensities of the pixels within the white area of the mask only (see Fig. 2).

To identify pixels that significantly differ in two sets of images, it is necessary to make images comparable. This can be done, e.g., by normalization of the spot intensities of individual gels to a constant sum. These two steps (background elimination and normalization) can be, however, replaced by a one step procedure, based on the regression method applied to the masked original images (i.e., images without background removal and without normalization). The main advantage of the proposed approach is that it can be performed automatically, i.e., without a need for optimization of input parameters of the background elimination method (intercept of the regression automatically accounts for this).

It ought to be pinpointed that the construction of the binary mask requires, however, a preliminary background removal and temporary normalization of all images. As the construction of a binary mask is not a crucial step in the entire procedure, its main goal being the elimination of the majority of irrelevant features (pixels) from further analysis, at this stage, background removal with the use of a rolling ball can be applied without fine tuning of input parameter (ball diameter). It is enough to define the diameter of the smallest spot to be detected. Once the binary mask is constructed, it is used to the original warped images (without background removal and without normalization).

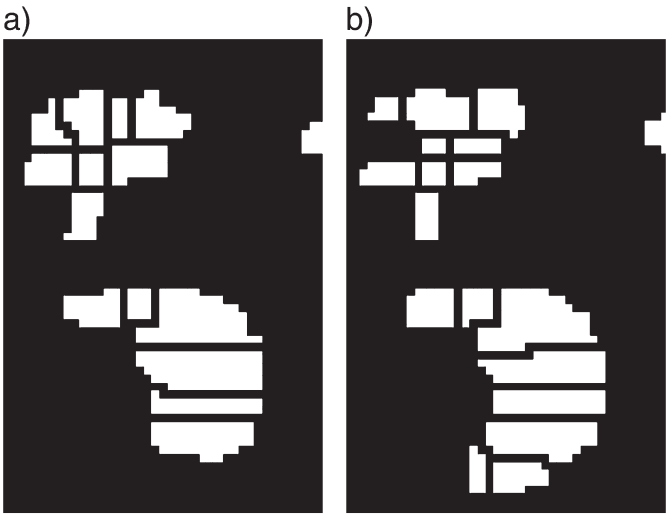


Fig. 1. Spots identified after background removal based on the rolling ball of the radius: a) 30 and b) 50.

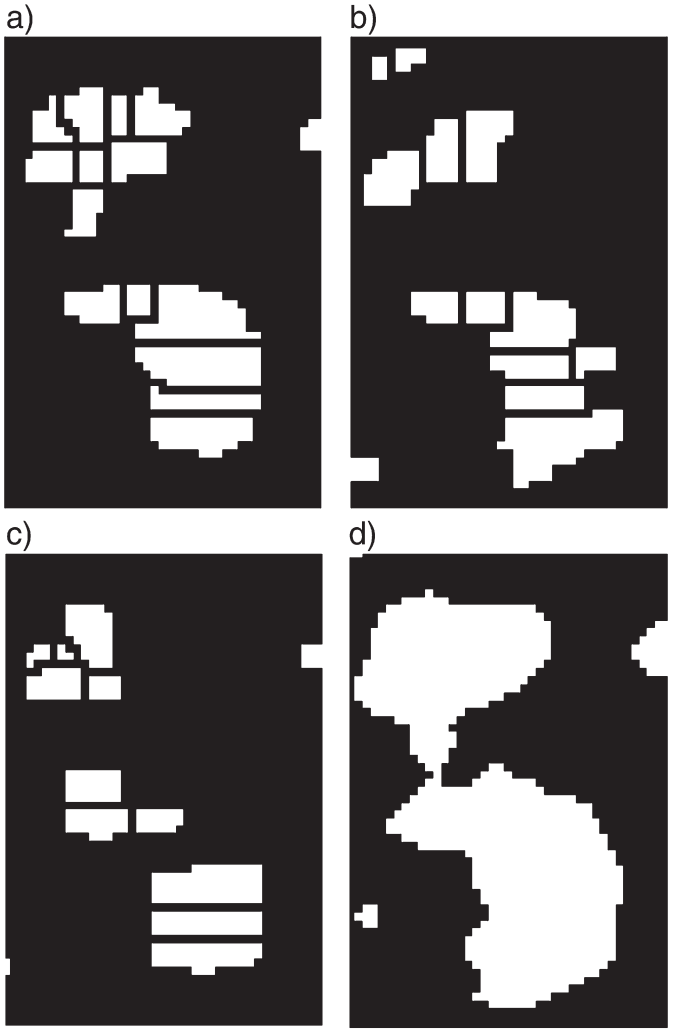


Fig. 2. Spots identified on the three different images (a–c) and the corresponding binary mask (d), constructed for the mean image of all studied images.

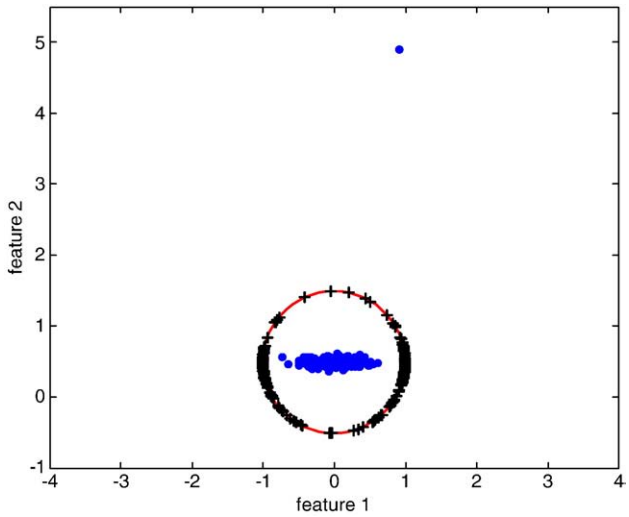


Fig. 3. Illustration of the principles of spherical Principal Component Analysis: all experimental points (blue) are projected on the sphere of unitary radius, placed in the robust center of the data. Their projections are presented as black crosses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

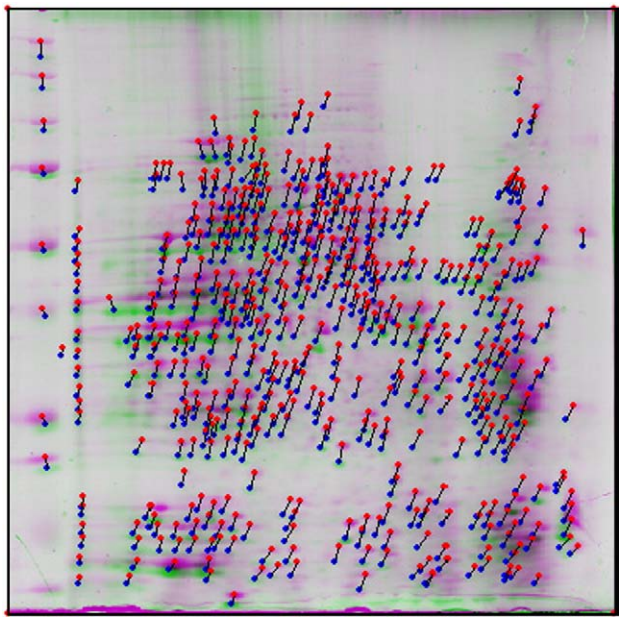


Fig. 4. Two overlapped images and the identified corresponding features.

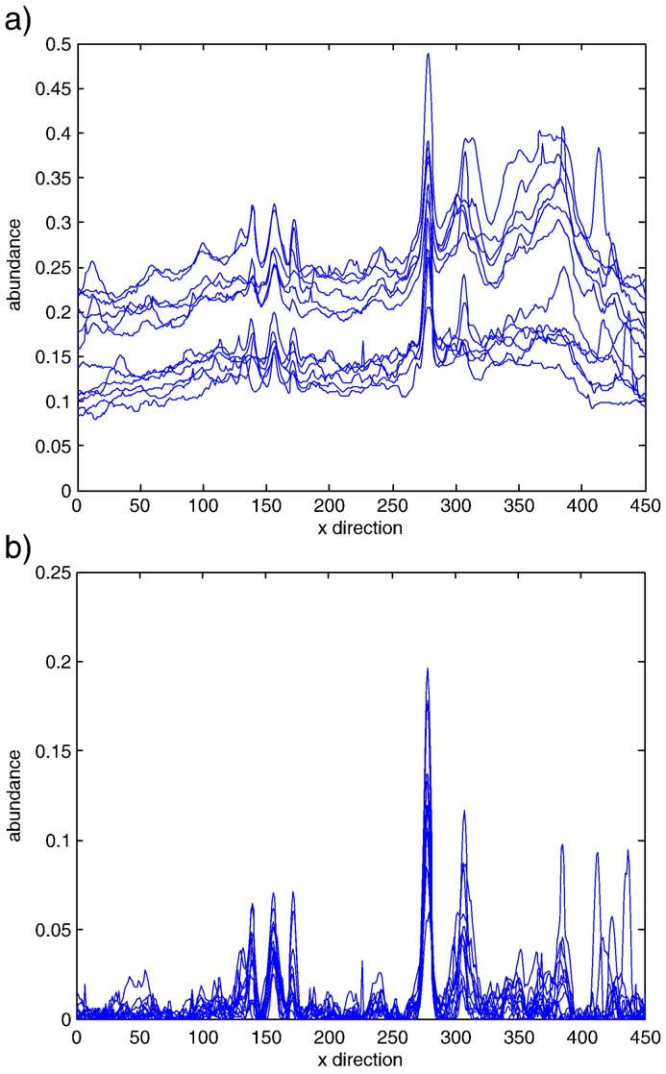


Fig. 5. Selected one-dimensional profiles of all studied images for: a) original images and b) after the background correction.

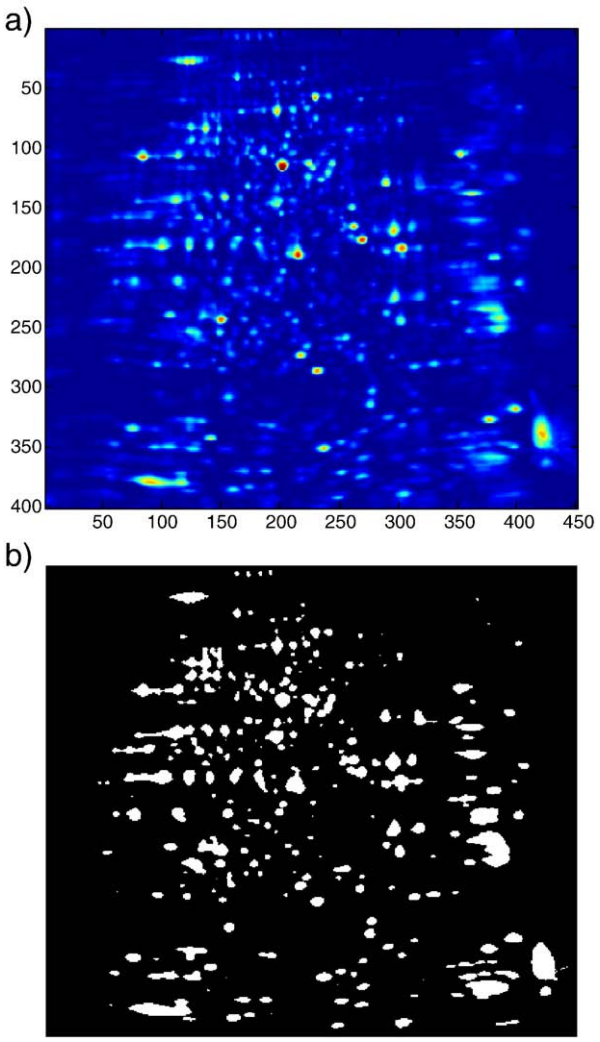


Fig. 6. Mean image and its binary mask for the studied data set.

2.2. Images preprocessing

As the reference image, either the mean image or the target image can be used. In our study, pixel intensities of the individual images are compared with the intensities of pixels of the mean image.

In an ideal case (no background, the same range of pixel intensities), the regression line of I_k versus I_{target} ought to have intercept equal to 0 and the unitary slope. In practice, the regression parameters can vary to a high extent. However, once their values are known it is possible to restore the ideal situation.

Knowing the parameters of regression for image k (and the target image), its intensities, I_k, can be transformed to I_k new as:

$$I_{k\ \text{new}} = (I_k - a)/b \tag{1}$$

where a and b denote, respectively, intercept and slope of the regression line.

As the regression method, Orthogonal Regression (OR) [26] was considered. In the case of data contamination, the use of robust versions of OR is of interest.

2.3. Orthogonal Regression and Robust Orthogonal Regression

Orthogonal Regression (or Total Least Squares) method [26] seems to be a straightforward choice to fit a linear model when there is no distinction between the independent and dependent variables. OR

minimizes the perpendicular distance from the data to the fitted model and is based on Principal Component Analysis (PCA).

The proposed robust version of orthogonal regression is based on the robust version of Principal Component Analysis, known as Spherical PCA [28,29]. The idea behind Spherical PCA is to project all objects on the hyper sphere of unitary radius with a center in the robust center of the data.

As shown in Fig. 3, the projected data points preserve the structure of the original data. However, the influence of the outlying objects is bounded by the down weighting them according to their distances from the robust data center defined by the L1-median.

Projecting the original data onto robust loadings, calculated for the down weighted objects, allows calculation of robust scores. While performing rOR, it is unnecessary to calculate robust scores, because regression coefficients are defined by the robust loadings only (see Appendix A).

The proposed rOR method is computationally simple and fast. The only problem can be associated with the calculation of the L1-median for the numerous data points. In such situations, it is recommended to replace L1-median with median.

2.4. Identification of significant features (pixels)

Warped and preprocessed images are then used for identification of proteins significantly differentiating the studied classes of samples. To this effect, many alternative approaches can be used. Among the popular univariate methods, there are those based on the correlation coefficient or the Fisher ratio. Using *t*-test, it is possible to calculate the probability, *p*, that a difference can occur by chance. However, due to a high number of simultaneously tested variables, there is a serious problem with the adjustment of the significance level for all the methods based on *t*-test. While testing the intensities of 20 000 pixels, 200 of them can be identified as significantly different for the two classes of studied samples just by chance at the significance level $p=0.01$. The Bonferroni correction [30], meant to deal with testing multiple parameters, is too conservative. In many cases, with its use no significant features can be found. In this study, we used the feature selection framework proposed in [1] and developed for significance analysis of microarray data. In this approach, the significance level is adjusted based on permutation of measurements and the false discovery rate (FDR) (percentage of pixels identified as significant by chance). The main idea of the discussed approach can be presented as follows:

1) Calculate the 'relative difference' $d(i)$ for the intensities of the ith pixel:

$$d(i) = \frac{\bar{x}_1(i) - \bar{x}_2(i)}{s(i) + s_0} \tag{2}$$

where $\bar{x}_1(i), \bar{x}_2(i)$ denote the average intensities of pixel i for class 1 and class 2, respectively; $s(i)$ denotes the standard deviation:

$$s(i) = \sqrt{a\left\{\sum_m \left[x_m(i) - \bar{x}_1(i)\right]^2 + \sum_n \left[x_n(i) - \bar{x}_2(i)\right]^2\right\}} \tag{3}$$

and

$$a = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{1}{(n_1 + n_2 - 2)} \tag{4}$$

where n_1, n_2 represent the number of samples in class 1 and class 2, respectively, whereas $m = 1, 2, \ldots, n_1$ and $n = 1, 2, \ldots, n_2$,

2) To eliminate the dependence of $d(i)$ on the pixel intensity, optimize the positive constant s_0 to minimize the coefficient of variation of $d(i)$ versus $s(i)$.
3) To find significant changes in feature (pixel) intensities, rank features according to their $d(i)$ values in the decreasing order and compare them to the expected relative differences $dE(i)$.Expected relative differences are calculated as the average of the predefined number of permutations, n_p: for each permuted data set, relative differences, $d_p(i)$ are calculated and ranked in the decreasing order, and $d_E(i)$ is calculated as:

$$d_E(i) = \frac{\sum_p d_p(i)}{n_p} \tag{5}$$

where i denotes the position of the relative differences in the ordered sequence of d for the given permutation, p.
4) Construct the scatter plot of $d(i)$ versus $d_E(i)$.
5) Find the optimal threshold value, th, to distinguish significant features from insignificant ones, by optimizing the false discovery rate (FDR), defined as:

$$FDR(\text{th}) = 100\frac{N_1(\text{th})}{N_2(\text{th})} \tag{6}$$

where N_1 holds for the number of falsely identified significant features (the permuted data) for a given threshold value and N_2 denotes the number of significant features in the experimental data set for the same threshold.

In our study, 1000 permutations were performed for each data set.

In the above approach, the 'relative difference' can be replaced by other parameters such as, e.g., the correlation coefficient.

2.5. Identification of significant features in multivariate approach

The approach presented so far is univariate, and thus it does not take into the account multivariate nature of the data studied. However, the idea of permutation test can be adapted to multivariate feature selection as well.

In our study, the Discriminant-Partial Least Squares (D-PLS) method was applied to model class belongingness of the samples studied. The multivariate feature selection procedure was based on the stability of regression coefficients [31]. For the studied data set (**X**, **y**), stability of regression coefficients can be calculated based on the leave-one-out cross-validation (CV) as:

$$\textbf{Stability} = \text{mean}(\mathbf{B})/\text{std}(\mathbf{B}) \tag{7}$$

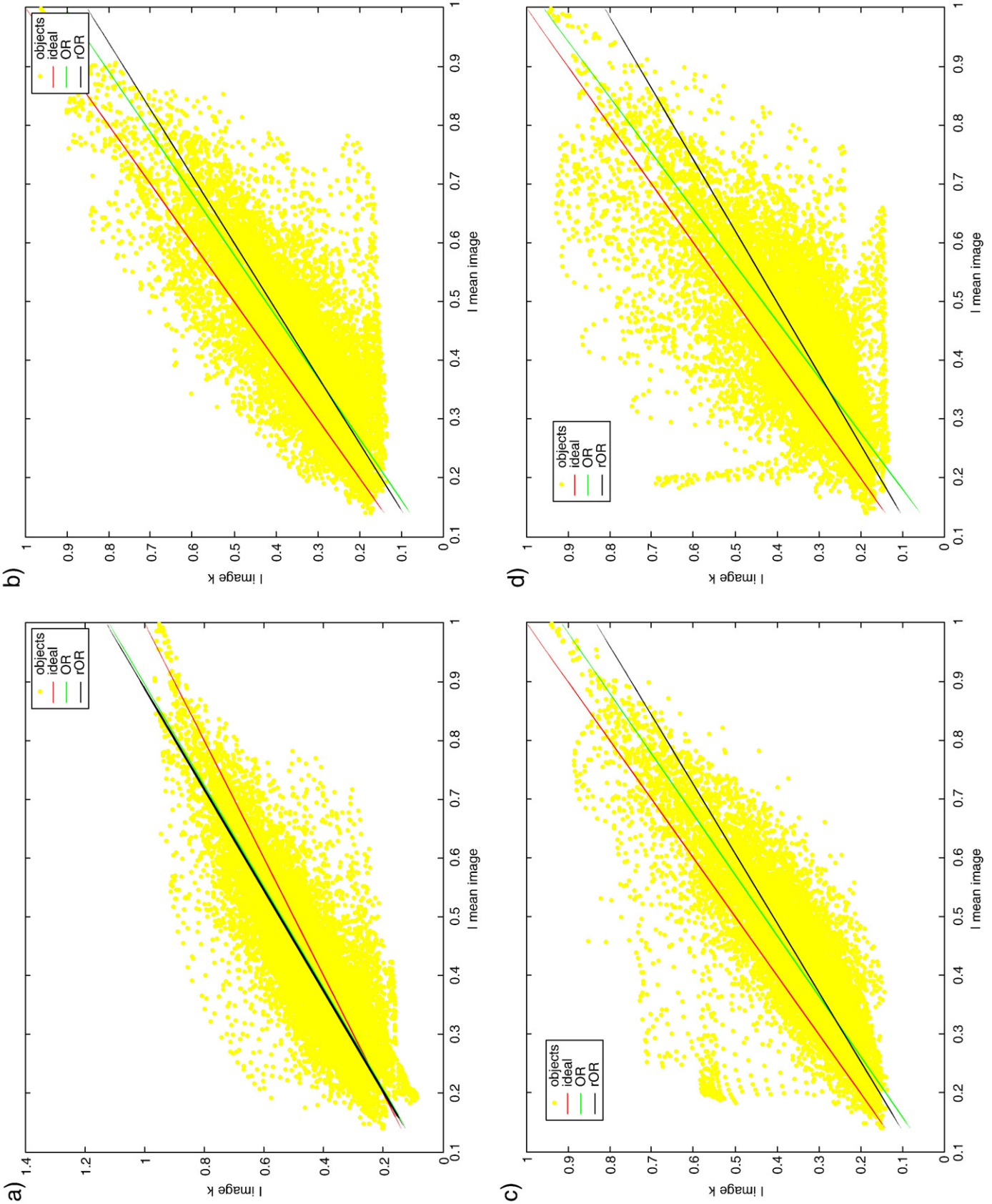
where rows of matrix **B** contain sets of regression coefficients calculated in the course of the CV procedure, and 'mean' and 'std' are row vectors containing means and standard deviations of the **B** columns.

Cut-off value (used to discriminate between stable and unstable coefficients) was calculated using permutation test, and namely: stability of regression coefficients was calculated 1000 times for the permuted data set, sorted in the increasing order and averaged. The cut-off value was adjusted to minimize FDR defined as a ratio of the number of stable variables from the permuted data to the number of stable variables from the original data set.

The variable selection procedure was leave-one-out model cross-validated to avoid model over-fitting. For the 12 studied samples there are 12 sets of the retained variables. Based on the frequency of occurrence of the individual variables in the retained sets, the final set of variables was determined. In this study, it contains all variables which appeared in more than 60% cases.

3. Data

Protein profiles studied in the current work represent a proteomics approach to the problem of repeated morphine administration and its


a)
b)
c)
d)
objects
ideal
OR
rOR
I image k
I mean image

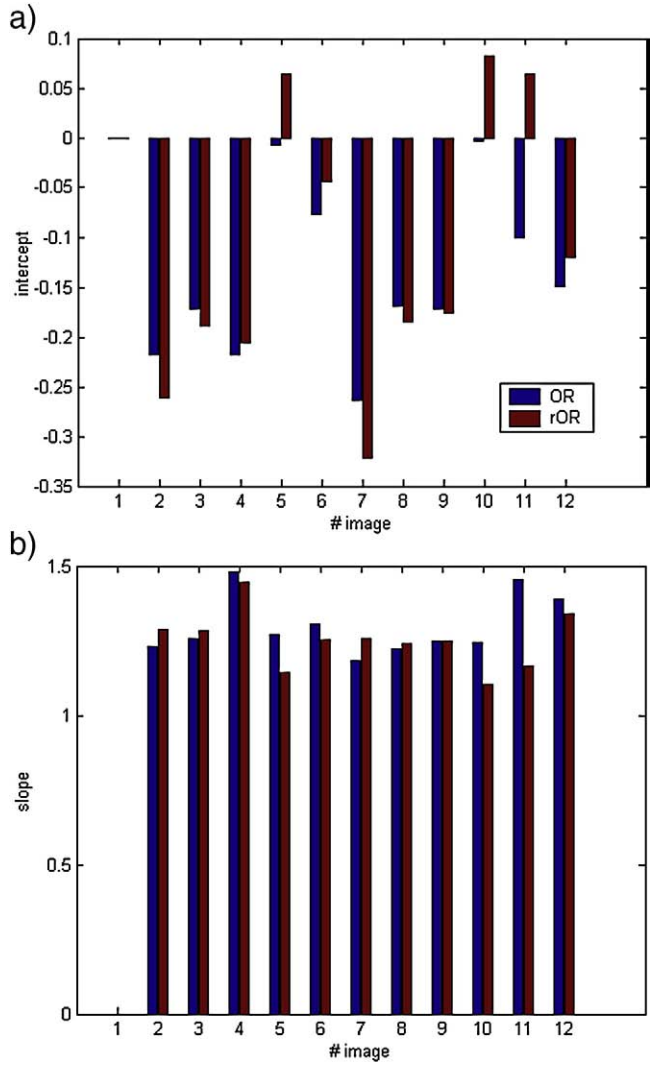


Fig. 8. Regression coefficients (a) intercepts and b) slopes) of the regression methods, OR and rOR.

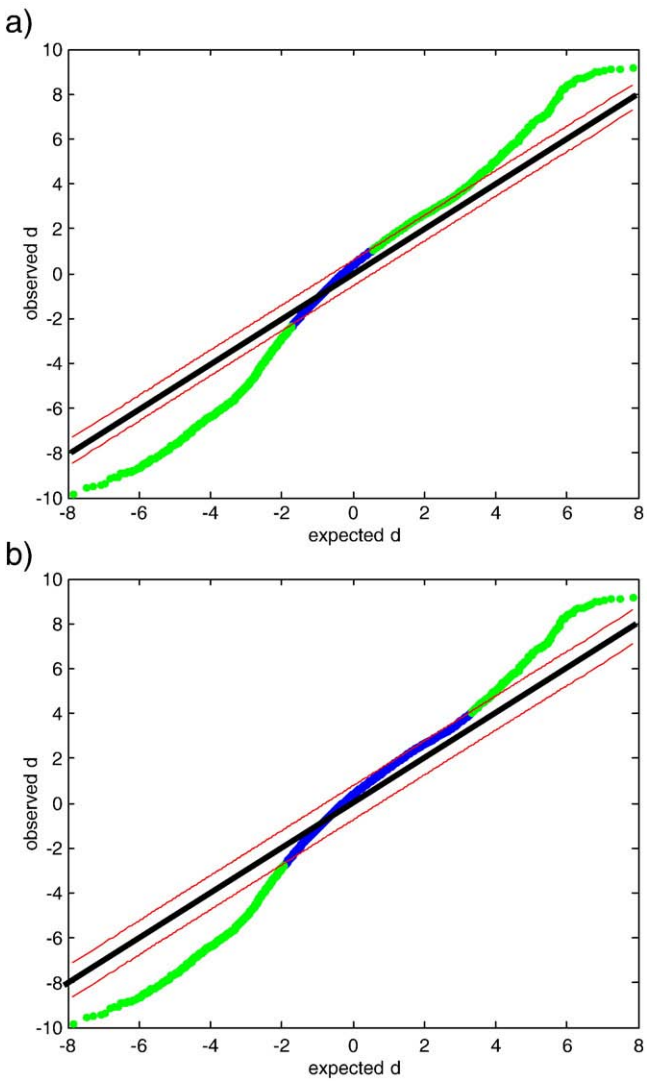


Fig. 9. Identification of significant features: scatter plot of the observed d (relative difference defined by Eq. (2)) versus expected d (defined by Eq. (5)) with the two consecutive bounds (thresholds). Black line indicates the line for observed d = expected d. Two red lines are drawn at the distance equal to the threshold value from solid line. Green points represent the significant and blue points represent the insignificant features for the given threshold value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

biochemical consequences. Opioids influence the way that neurons work and communicate in the central nervous system, thus generating adaptive changes in intracellular mechanisms. Such variations, represented by qualitative and quantitative changes in the phenotype, can be monitored by comparison with control samples.

Quantitative analysis of two-dimensional gel sets, performed for striatum of morphine-administered rats and controls, resulted in identification of morphine-altered protein expression. In total, six striata from control, and six from morphine treated rats were analyzed. Sample preparation procedures and separation conditions and are described in detail in [32].

4. Results and discussion

At the initial stage of data analysis, the twelve studied images, representing two classes of objects, were warped to image 1, using the fuzzy warping method [9]. The target image was selected as the one with the highest correlation coefficient with respect to all other images [33]. As a global transform function, the second-order polynomial was applied, and once the corresponding features were identified (see Fig. 4), the piece-wise linear transform of the image grid was performed. Then, the 'bicubic' method was used for interpolation of image intensities.

As shown in Fig. 5, background component significantly contributes to the overall signal and it differs from one sample to another.

Regression methods can be used to remove background (intercept) and to rescale the studied signals (slope) simultaneously (see Fig. 5b). It ought to be pinpointed that the regression approaches are applied to the data points from white pixels of the binary mask (see Fig. 6).

As the binary mask contains 12 343 white pixels, the final data sets has the dimensionality $12 \times 12\,343$. The two studied regression methods, orthogonal least squares (OR) and its robust version (rOR), were applied to $\mathbf{X}$ data. The differences between the studied regression methods are illustrated in Fig. 7.

Fig. 7. Regression models built for four selected images and mean images; each subplot represents the linear orthogonal regression (OR) and robust orthogonal regression (rOR) fits to the scatter plot of the points representing abundance of pixels in the kth image and the mean image. In an ideal case, when no correction is needed, regression line with slope equal to 1 would represent linear fit to the scatter plot.

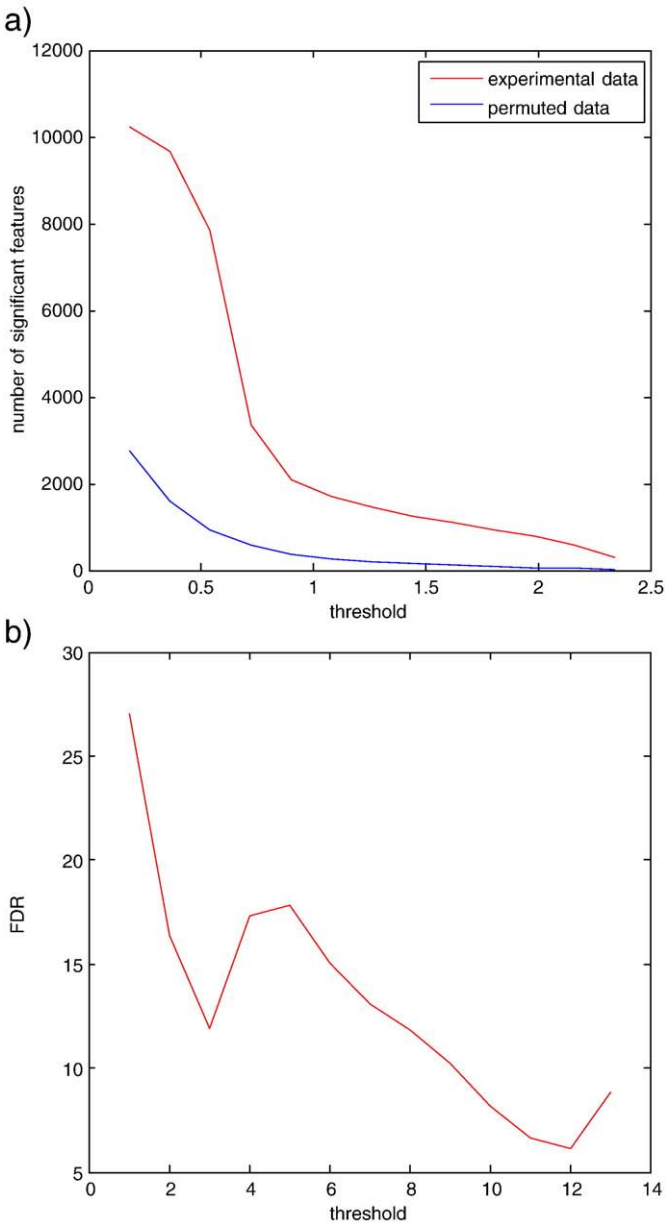


Fig. 10. a) Number of significant features versus the threshold value for the experimental data and permuted data, and b) false discovery rate versus the threshold value.

For instance, for the first image, presented in Fig. 7a, there are no difference between the OR and rOR methods. Differences between OR and its robust version are observed for images presented in Fig. 7b–d.

In Fig. 8, the regression coefficients (intercept and slope), calculated for all 12 images are demonstrated for the studied regression methods.

In the chemometric community, the use of LS or OR for data preprocessing is quite popular (e.g., [34–39]), but up to our best knowledge, their robust versions are not. For processing of the proteomic data, robust version of OR seems to be particularly interesting. In OR, both dependent and independent variables are assumed to have similar measurement errors, which is the case in this study, where individual images are compared with the target image.

Moreover, many local distortions of gel images cannot always efficiently be eliminated and thus can lead to the presence of outlying elements in the data matrix. The OR method is based on the Principal Component Analysis and thus it can be influenced by the outlying objects. However, its robust version can be resistant to any type of outlying objects. Of course, not always is there a need for the robust version of OR, but it can be applied as a standard method, because it is efficient, i.e., for the non-contaminated data it approaches the solution of OR.

Images corrected using the discussed regression methods were compared based on FDR. For each data set, the value of s_0 was selected to stabilize the variation of $d(i)$ versus $s(i)$, and the permutation test was performed to choose the bounds of the scatter plot of the observed relative differences $d(i)$ versus the expected relative differences $d_E(i)$ minimizing FDR. For an illustrative example, two consecutive bounds are presented in Fig. 9.

Fig. 10a presents dependence of the number of significant features (pixels) for experimental and permuted data sets on the threshold value and in Fig. 10b, the FDR versus the threshold value is shown.

The threshold value, for which FDR reaches the minimum, is selected as the optimal one and the identified significant features can be visualized on the mean image.

OR and rOR lead to the FDR values equal to 3.9 and 3.0, respectively. Lower value of FDR is observed for rOR and it is comparable with the results obtained for data preprocessed using penalized asymmetric least squares method of background elimination.

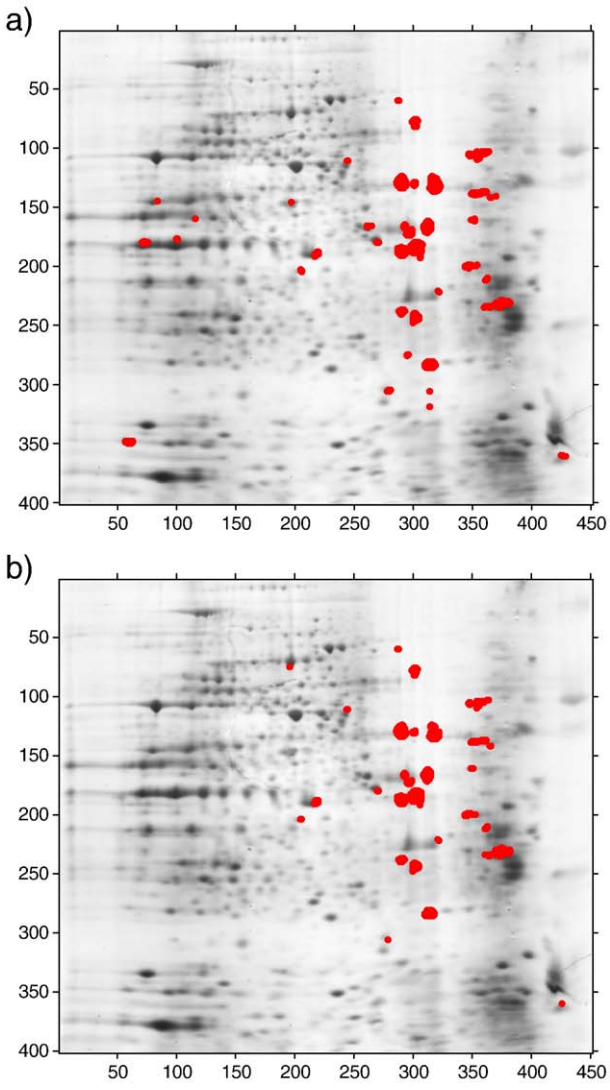


Fig. 11. Mean image with marked significant pixels identified with the OR and rOR preprocessing methods, respectively.

The number of the significant pixels varies from 743 (rOR) to 497 (OR), but the set of the spots including these pixels is very similar (see Fig. 11).

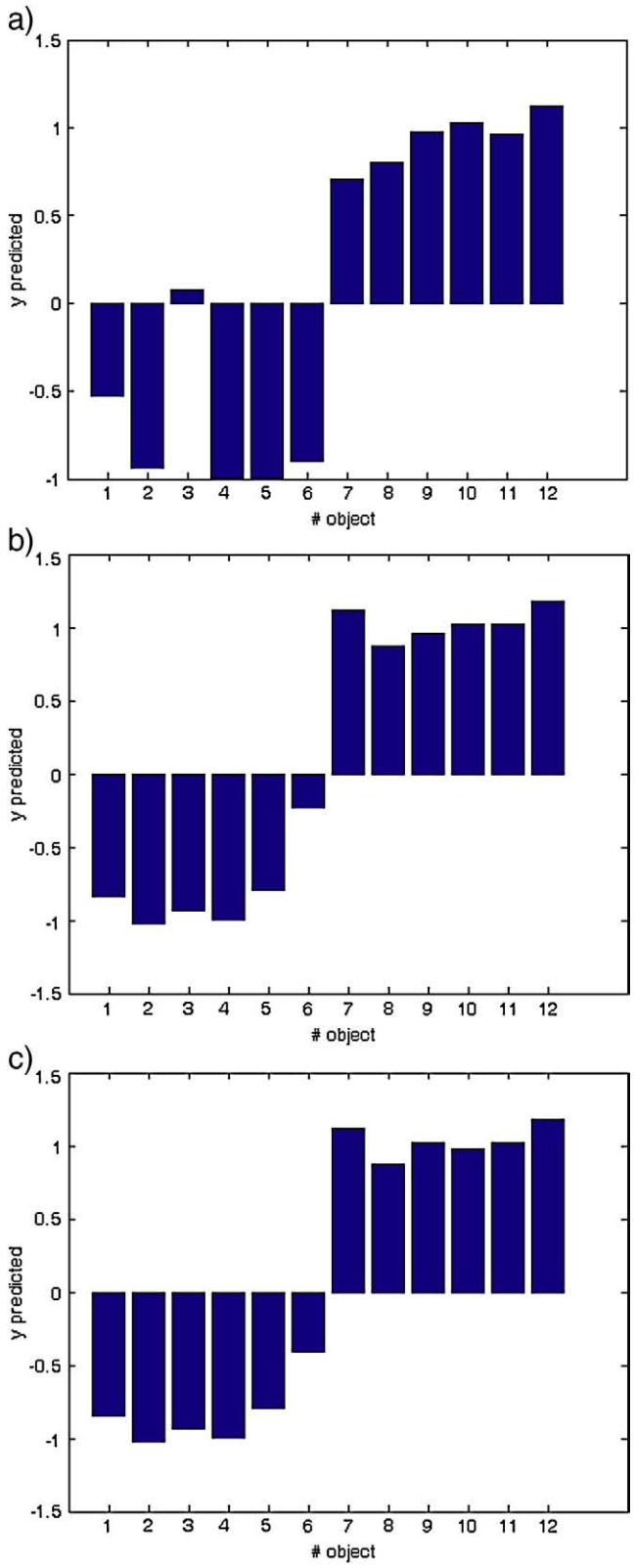


Fig. 12. The predicted *y* values for the studied data set, preprocessed using: a) OR; b) rOR and c) PALS methods. Negative values of *y* indicate belongingness of samples to class 1, whereas positive values of *y* indicate belongingness to class 2.

Although in this study we focus on the univariate evaluation of pixels significance, it can also be done based on a multivariate approach such, as e.g., Partial Least Squares (PLS) with embedded feature selection procedure [40,41]. The multivariate feature selection approach is of particular interest, when no individual spots offer a significant discrimination power, but the data still contain information sufficient for medical diagnostic.

Using D-PLS, we do not expect selected features to be the same as the features identified based on 'relative distance' parameter, because their selection highly depends on the applied objective function. In D-PLS, we identify the features that are highly correlated with *y* and simultaneously well describing the variance in **X**. The main purpose of our study was to estimate the influence of data preprocessing on the final results of data analysis. As demonstrated in Fig. 12, the Root Mean Square Error of model-CV (RMS-MCV), obtained from the data preprocessed with rOR is comparable with the error observed, when the penalized asymmetric least squares method is applied for background elimination.

5. Conclusions

The proposed robust orthogonal regression method can be effectively used for preprocessing of the 2D gel masked images. It does not require optimization of input parameters, so its implementation to the start-to-end approach in fact makes the whole approach automatic.

Acknowledgment

This work was supported by the International Centre for Genetic Engineering and Biotechnology (ICGEB) grant no. CRP/POL05-02.

Appendix A

Matlab code of Orthogonal Regression

```
function [b] = OR(x,y);
% INPUT:    x and y – column vectors of independent and dependent
%           variable, respectively
% OUTPUT:   b – vector of regression coefficients
%           b = [intercept, slope]

b = zeros(2,1);
[m,n] = size(x);
X = [x y];
mX = mean(X);
Xc = X-ones(m,1)*mX;
[s,v,d] = svd2(Xc);
b(2,1) = d(2,1)/d(1,1);
b(1,1) = mX(1,2)-b(2)*mX(1,1);

function [U,S,V] = svd2(X)
[m,n] = size(X);
if m<n              % If more variables than objects, use the kernel version
   [U,S] = eig(X*X');
   S = diag(S);
   [a,b] = sort(S);
   b = flipud(b);
   S = S(b);
   S = abs(S);
   U = U(:,b);
   V = X'*U*diag(sqrt(1./S));
   S = diag(sqrt(S));
else                % Else, normal SVD algorithm
  [U,S,V] = svd(X,0);
end
```

Matlab code of robust Orthogonal Regression

```
function [b] = robustOR(x,y);
% INPUT:   x and y – column vectors of independent and dependent
%          variable, respectively
% OUTPUT:  b – vector of regression coefficients [intercept, slope]

b = zeros(2,1);
[m,n] = size(x);
X=[x y];
mX = median(X);
Xc = X-ones(m,1)*mX;
[d] = sPCA(Xc);
b(2,1) = d(2,1)/d(1,1);
b(1,1) = mX(1,2)-b(2)*mX(1,1);
```

Matlab code of spherical PCA

```
function d = sPCA(X);

% INPUT:   X - data matrix of size (m,n) containing m samples and
           n variables
% OUTPUT: d – robust loadings

[m,n] = size(X);
th = median(X);        % or L1median(X) if m is small
X = X-ones(m,1)*th;
w = ones(m,1)./(sqrt(sum((X.^2)')))';
X = ((w*ones(1,n)).*X)./sum(w(:,1));
[s,v,d] = svd2(X);
```

References

[1] V. Goss Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, P. Natl. Acad. Sci. U. S. A. 98 (2001) 5116–5121.
[2] A.W. Wheelock, A.R. Buckpitt, Software-induced variance in two-dimensional gel electrophoresis image analysis, Electrophoresis 26 (2005) 4508–4520.
[3] N. Campostrini, L.B. Areces, J. Rappsilber, M.C. Pietrogrande, F. Dondi, F. Pastorino, M. Ponzoni, P.G. Righetti, Spot overlapping in two-dimensional maps: a serious problem ignored for much too long, Proteomics 5 (2005) 2385–2395.
[4] H. Grove, K. Hollung, A.K. Uhlen, H. Martens, E.M. Færgestad, Challenges related to analysis of protein spot volumes from two-dimensional gel electrophoresis as revealed by replicate gels, J. Proteome Res. 5 (2006) 3399–3410.
[5] M. Daszykowski, I. Stanimirova, A. Bodzon-Kulakowska, J. Silberring, G. Lubec, B. Walczak, The start-to-end processing of two-dimensional gel electrophoretic images, J. Chromatogr. A 1158 (2007) 306–317.
[6] B. Walczak, W. Wu, Fuzzy warping of chromatograms, Chemom. Intell. Lab. Syst. 77 (2005) 173–180.
[7] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Baseline reduction in two dimensional gel electrophoresis images, Acta Chromatogr. 15 (2005) 82–96.
[8] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Preprocessing of two-dimensional gel electrophoresis images, Proteomics 4 (2004) 2377–2389.
[9] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Matching of 2D gel electrophoresis images, J. Chem. Inf. Comp. Sci. 43 (2003) 978–986.
[10] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Feature based fuzzy matching of 2D gel electrophoresis images, J. Chem. Inf. Comp. Sci. 42 (2002) 1431–1442.
[11] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Comparison of image transformations methods used in matching of 2D gel electrophoresis images, Acta Chromatogr. 13 (2003) 7–21.
[12] E.M. Færgestad, M. Rye, B. Walczak, L. Gidskehaug, J.P. Wold, H. Grove, X. Jia, K. Hollung, U.G. Indahl, F. Westad, F. van den Berg, H. Martens, Pixel-based analysis of multiple images for the identification of changes: a novel approach applied to unravel proteome patters of 2-D electrophoresis gel images, Proteomics 7 (2007) 3450–3461.
[13] P.H.C. Eilers, I.D. Currie, M. Durban, Fast and compact smoothing on large multidimensional grids, Comput. Stat. Data Anal. 50 (2006) 61–76.
[14] S.R. Sternberg, Biomedical image processing, IEEE Computer 16 (1983) 22–34.
[15] P.F. Lemkin, C. Merril, L. Lipkin, M. Van Keuren, W. Oertel, B. Shapiro, M. Wade, M. Schultz, E. Smith, Software aids for the analysis of 2D gel electrophoresis images, Comput. Biomed. Res. 12 (1979) 517–544.
[16] P.F. Lemkin, L.E. Lipkin, E.P. Lester, Some extensions to the GELLAB 2D electrophoresis gel analysis system, Clin. Chem. 28 (1982) 840–849.
[17] J.I. Garrels, The QUEST system for quantitative analysis of two-dimensional gels, J. Biol. Chem. 264 (1989) 5269–5282.
[18] P.F. Lemkin, Comparing two-dimensional gels across the Internet, Electrophoresis 18 (1997) 461–470.
[19] R.D. Appel, J. Vargas, P.M. Palagi, D. Walter, D.F. Hochstrasser, Melanie II – a third-generation software package for analysis of two-dimensional electrophoresis images: II, Algorithms, Electrophoresis 18 (1997) 2735–2748.
[20] K.P. Pleissner, F. Hoffmann, K. Kriegel, C. Wenk, S. Wegner, A. Sahlström, H. Oswald, H. Alt, E. Fleck, New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases, Electrophoresis 20 (1997) 755–765.
[21] F. Hoffmann, K. Kriegel, C. Wenk, An applied point pattern matching problem: comparing 2D patterns of protein spots, Discrete Appl. Math. 93 (1999) 75–88.
[22] J. Pánek, J. Vohradský, Point pattern matching in the analysis of two-dimensional gel electropherograms, Electrophoresis 20 (1997) 3484–3491.
[23] S. Veeser, M.J. Dunn, G.-Z. Yang, Multiresolution image registration for two-dimensional gel electrophoresis, Proteomics 1 (2001) 856–870.
[24] Z. Smilansky, Automatic registration for images of two-dimensional protein gels, Electrophoresis 22 (2001) 1616–1626.
[25] K. Conradsen, J. Pedersen, Analysis of 2-dimensional electrophoretic gels, Biometrics 48 (1992) 1273–1287.
[26] S. Van Huffel, J. Vandewalle, The Total Least Squares Problem: Computational Aspects and Analysis, SIAM, Philadelphia, 1991.
[27] G.R. Phillips, E.M. Eyring, Comparison of conventional and robust regression in analysis of chemical data, Anal. Chem. 55 (1983) 1134–1138.
[28] N. Locantore, J.S. Marron, D.G. Simpson, N. Tripoli, J.T. Zhang, K.L. Cohen, Robust principal component analysis for functional data, Test 8 (1999) 1–73.
[29] M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, B. Walczak, Robust statistics in data analysis – a review. Basic concepts, Chemom. Intel. Lab. Syst. 85 (2007) 203–219.
[30] E.M. Weisstein, "Bonferroni Correction" MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/BonferroniCorrection.html.
[31] V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B. Vandeginste, C. Sterna, Elimination of uninformative variables for multivariate calibration, Anal. Chem. 68 (1996) 3851–3858.
[32] A. Bierczynska-Krzysik, J.P. Pradeep, J. Silberring, J. Kotlinska, T. Dylag, M. Cabatic, G. Lubec, Proteomic analysis of rat cerebral cortex, hippocampus and striatum after exposure to morphine, Int. J. Mol. Med. 18 (2006) 775–784.
[33] M. Daszykowski, B. Walczak, Target selection for alignment of chromatographic signals obtained using monochannel detectors, J. Chromatogr. A 1176 (2007) 1–11.
[34] H. Martens, T. Næs, Multivariate Calibration, John Wiley & Sons, Chichester, UK, 1991.
[35] T.K. Karakach, R.M. Flight, P.D. Wentzell, Bootstrap method for the estimation of measurements uncertainty in spotted dual-color DNA microarrays, Anal. Bioanal. Chem. 389 (2007) 2125–2141.
[36] I.S. Helland, T. Næs, T. Isaksson, Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data, Chemom. Intel. Lab. Syst. 29 (1995) 233–241.
[37] P. Geladi, D. MacDougall, H. Martens, Linearization and scatter-correction for near-infrared reflectance spectra of meat, Appl. Spectrosc. 39 (1985) 491–500.
[38] T. Isaksson, T. Naes, The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy, Appl. Spectrosc. 42 (1988) 1273–1284.
[39] J.L. Ilari, M. Martens, T. Isaksson, Determination of particle size in powders by scatter correction in diffuse near-infrared reflectance, Appl. Spectrosc. 42 (1988) 722–728.
[40] T. Czekaj, W. Wu, B. Walczak, Classification of genomic data: some aspects of feature selection, Talanta 76 (2008) 564–574.
[41] L. Gidskehaug, E. Andersen, B.K. Alsberg, Cross model validation and optimisation of bilinear regression models, Chemom. Intell. Lab. Syst. 93 (2008) 1–10.