Available online at www.sciencedirect.com

ScienceDirect

Analytica Chimica Acta 595 (2007) 299–309

ANALYTICA CHIMICA ACTA

www.elsevier.com/locate/aca

Visualisation and interpretation of Support Vector Regression models

B. Üstün, W.J. Melssen, L.M.C. Buydens *

Institute for Molecules and Materials, Analytical Chemistry, Radboud University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Received 9 October 2006; received in revised form 1 March 2007; accepted 2 March 2007
Available online 18 March 2007

Abstract

This paper introduces a technique to visualise the information content of the kernel matrix and a way to interpret the ingredients of the Support Vector Regression (SVR) model. Recently, the use of Support Vector Machines (SVM) for solving classification (SVC) and regression (SVR) problems has increased substantially in the field of chemistry and chemometrics. This is mainly due to its high generalisation performance and its ability to model non-linear relationships in a unique and global manner. Modeling of non-linear relationships will be enabled by applying a kernel function. The kernel function transforms the input data, usually non-linearly related to the associated output property, into a high dimensional feature space where the non-linear relationship can be represented in a linear form. Usually, SVMs are applied as a black box technique. Hence, the model cannot be interpreted like, e.g., Partial Least Squares (PLS). For example, the PLS scores and loadings make it possible to visualise and understand the driving force behind the optimal PLS machinery.

In this study, we have investigated the possibilities to visualise and interpret the SVM model. Here, we exclusively have focused on Support Vector Regression to demonstrate these visualisation and interpretation techniques. Our observations show that we are now able to turn a SVR black box model into a transparent and interpretable regression modeling technique.

Keywords: Support Vector Regression; Feature space; Kernel functions; Non-linear regression; Model visualisation and interpretation

1. Introduction

Partial Least Squares (PLS), one of the commonly used regression techniques, has the possibility to interpret the obtained regression model by relating the results to the original input space by the loadings and regression coefficients [1]. The PLS loadings indicate which input variables (e.g., wavelengths or frequencies in case of spectral data) are highly correlated to the predicted output values.

Support Vector Machines (SVMs) [2,3] become increasingly popular for solving classification as well as regression problems, largely stimulated by its possibility to model data possessing non-linear relationships by employing the so-called kernel trick [4]. This kernel transformation maps the original input space into a high dimensional feature space where the non-linear relationship can be modeled in a linear way. The transformation to this feature space is accomplished by using a specific kernel function. Several kernel functions like, variance–covariance based linear and polynomial kernels, the Euclidean distance based Radial Basis Function (RBF) and the Pearson VII Universal Kernel (PUK) functions are proposed in literature [4–6]. However, due the fact that SVMs employ a kernel it has the drawback that it will lose the correlation between the obtained SVM model and the original input space. In other words, which part(s) of the input data (e.g., which wavelength region(s)) contribute to the SVM results. The kernel transforms the original input space with the dimension ($N \times M$) into a feature space with dimension ($N \times N$), where N is the number of objects (samples) and M the number of variables (e.g., wavelengths). So, the input matrix will be transformed into a square matrix, called the kernel matrix. By this transformation, the information regarding the original input variables is vanished and a direct interpretation (like in PLS) of the SVM model is not possible. A lot of papers studying SVMs, report just a high performance for classification and regression problems, but unfortunately do not comment on the underlying relationship between the input and modeled output data. Hence, in these papers SVMs are used as black boxes.

In this paper, we present an approach to visualise and interpret the SVM model. This approach gives us the opportunity to answer the question: what is happening in the complex SVM

[*] Corresponding author. Tel.: +31 24 36 53192; fax: +31 24 36 52653.
E-mail address: L.Buydens@science.ru.nl (L.M.C. Buydens).

0003-2670/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2007.03.023

machinery and which variables of the input space are the most important ones. This type of analysis will turn SVMs into a transparent and understandable modeling technique.

Here, we will focus solely on (non-)linear regression problems (function approximation) and apply the Support Vector Regression (SVR) methodology [4,5] to demonstrate how the results can be visualised and interpreted. One simulated data set and three real-world spectral data sets were selected for this study. The simulated data set is used to illustrate the applicability of our proposed visualisation and interpretation method. The analysis method is tested on the three real-world data sets.

The outline of the paper is as follows. Section 2 concentrates on the concept of our approach to visualise and interpret the SVR results. The data specification and the generalisation criterion are given in Section 3. The experimental results and discussion are presented in Section 4. Some final remarks and overall conclusions are given in Section 5.

2. Theory

This Section describes the basic concept of our approach to visualise and interpret the SVR model. The theory of Support Vector Regression models has been extensively described in literature [4,5,7]. Therefore, only a brief description of the concept of SVR will be given in Section 2.1, while Section 2.2 concentrates on the aspects of visualisation and interpretation of SVR models on basis of a simulated data set.

2.1. Support Vector Regression

The basic idea behind SVR is that the original input space, which is usually non-linearly related to the predictor variable, is mapped onto a higher dimensional feature space by means of a non-linear mapping function (kernel function) to deal with these non-linearities. The feature space, which is explicitly embedded in the kernel matrix, will be used as a new input set to solve the regression problem.

Consider a data set **X** ($N \times M$) with an output vector $\boldsymbol{y}_i \in R$. The objective of SVR is to find a multivariate regression function $f(\mathbf{x})$ based on the given data set S to predict the desired output property (e.g., the concentration of a chemical compound) of an unknown object (e.g., a spectrum). The complete SVR equations are fully described in [2,4,5] and will not be repeated here. Therefore, we will immediately write down the SVR equation as follow:

$$f(\mathbf{x}) = \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*) \left\langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \right\rangle + b \tag{1}$$

where α_i and α_i^* represent the Lagrange multipliers satisfying the constraint $0 \le \alpha_i, \alpha_i^* \le C$ [5]. C is a regularisation constant which determines the trade-off between the training error and model simplicity. A more detailed description of Eq. (1) and the parameters α and C are given in [4,5]. The entity ϕ represents the mapping function and the parameter b the offset of the regression function $f(\mathbf{x})$. Since, the non-linear mapping $\phi(\mathrm{x})$ is in general unknown on beforehand and therefore difficult to determine. The mapping term $\langle \phi(\mathrm{x}_i)\ \phi(\mathrm{x}_j) \rangle$ in Eq. (1) is often approximated by a kernel function:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \left\langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \right\rangle \tag{2}$$

This kernel transformation allows us to handle non-linear relationships in the data in an easier way. Several kernel functions, such as variance–covariance, polynomial, Radial Basis Function (RBF) based kernels and the Pearson VII Universal Kernel (PUK) are suited to tackle (non-)linear regression problems. In our previous publication, we have shown that the PUK kernel function is capable to serve as a generic alternative to the commonly applied variance–covariance, polynomial and RBF based kernel functions [6]. For this reason, we will use the PUK function as being the generic one throughout this paper.

After invoking the kernel function, Eq. (1) becomes equal to:

$$f(\mathbf{x}) = \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*) \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + b \tag{3}$$

As mentioned before, the kernel function usually transforms the non-linear input space into a high-dimensional feature space in which the solution of the problem can be represented as being a straight linear problem. The α_i and α_i^* values are determined by solving the transformed regression problem by means of Quadratic Programming (QP) [3,8]. Only the input objects corresponding to the non-zero Lagrange Multipliers α_i and α_i^* do contribute to the final regression model and therefore are named support vectors. A detailed description including the properties and advantages of support vectors is given in [4,5].

2.2. Visualisation and interpretation

The mapping of the input data into a higher dimensional feature space by using a kernel function results in a weighted similarity matrix. The weighted similarity matrix, hereafter named kernel matrix, represents the similarity between the objects of the training set. The kernel matrix is square and symmetrical, and its number of rows and columns corresponds to the number of objects in the training set. By transforming the input data into a kernel matrix the information about the original input variables (e.g., wavelengths in case of spectral data) is lost. A drawback of this is that a direct interpretation of the final SVR model in relation to the involved input variables is not possible. To circumvent this problem, we have developed a methodology to visualise the information embedded in the kernel matrix and to interpret the optimised SVR model. Visualisation will makes clear which part(s) of the original data is used or plays a more important role for the transformation of the non-linear problem into a linear problem. In this way, it becomes clear to understand and to present the influence/effect of the kernel transformation. I will probably help to understand the positive contribution of kernels in case of SVR but also in other kernel-based techniques. Why are kernel-based methods successful? The main task of the interpretation method is to determine the responsible part(s) of the kernel matrix, which can be related to the original data, to the final SVR model. Except the determination of the part(s) which are responsible for the nice results of SVR it has also the inten-

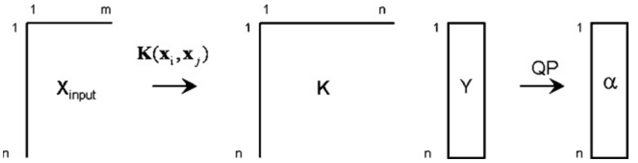

Fig. 1. In SVR, the original non-linear input space $\mathbf{X}_{\text{input}}$ will be transformed into a higher-dimensional feature space represented by a symmetric square matrix **K**, called kernel matrix, by using a kernel function $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$. In this way, the non-linear problem becomes linear problem in the feature space. The kernel matrix will be used as input together with the accompanying output values Y to solve the regression problem by Quadratic Programming (QP). The latter step will resulting **α**-vector. The non-zero α_i-values are the so-called support vectors.

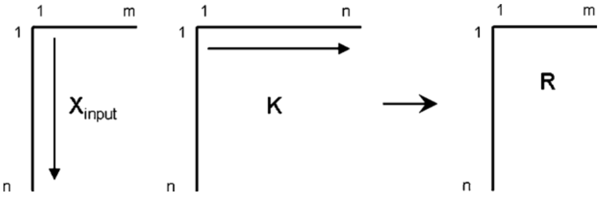

Fig. 2. A correlation matrix R is created by calculating the correlation between each column of the input data with that of each row of the kernel matrix. A column of the input data consist the input variables (e.g., spectral variables in case of spectral data) for all objects, while each row of the kernel matrix represents a similarity measure of a specific object with all other objects.

tion to understand the power of SVR. What is the driving force behind SVR to build a model on base of the whole or part(s) of information extracted by the kernel?

The mentioned visualisation and interpretation analysis methods are described in the following sections. To make the basics of these methods understandable, an overview of the different steps of the SVR algorithm is given by Fig. 1.

2.2.1. Visualisation of the information in the kernel matrix

The first step in SVR is mapping of the input data by using a kernel function into a kernel matrix (Fig. 1). The main objective of this mapping is to linearise the problem: e.g., the non-linear regression problem will be transformed into a high dimensional space in which the solution of the problem can be represented as being a straight linear problem. It is very important that the non-linear problem will be transformed into a linear problem without information loss about the desired property. The choice of the kernel function and especially its parameter settings are of utmost importance to solve the problem adequately. If the kernel parameters of the selected kernel function are not chosen properly, valuable information will be lost and SVR, as well other kernel-based techniques, will fail to solve the regression problem. Thus, it is a crucial step to select the optimal kernel function and its accompanying parameters to describe the nature of the original input data. In this paper, we apply the generic PUK [6] kernel function; its parameters are optimised by a Genetic Algorithm (GA) followed by a Simplex optimisation procedure [9].

A kernel defines a square and symmetrical matrix, which represents the similarity between pairs of objects. Each row (and column) in the kernel matrix represents the similarity of a specific object with all other objects in the training set. Since the PUK function is used here, the element values of the kernel matrix will vary between zero and one. A value close to zero indicates that two objects are very different, whereas a value close to one corresponds to two almost similar objects. As mentioned earlier, a drawback of this representation is the loss of information about the input variables. In other words, by mapping the original data into a kernel matrix the information related to the original input variables is lost. This information might be relevant to interpret which input variables (e.g., wavelength(s)) are responsible for the constructed regression model.

To relate the information of the kernel matrix back to the input variable space the correlation between each column (variable) of the original input data matrix (**X**) and each row of the kernel matrix, **K**, is calculated (Fig. 2). Since each column of **X** contains the information per input variable and each row of **K** represents the similarity between the objects, the elements of the calculated correlation matrix (**R** ($M \times N$)) will express the contribution of each input variable to the kernel matrix. A correlation value close to zero indicates that the respective input variable has no relevant contribution to a specific row of the kernel matrix, while a value close to +1 or −1 indicates that it is an important variable. An image plot of the correlation matrix, hereafter named Correlation Image (CI), makes it possible to visualise the importance of the input variables with respect to the kernel matrix. In this way, it becomes possible to visualise the information contained in the kernel matrix.

In order to demonstrate the kernel visualisation method, we will apply it to a two component regression problem (Fig. 3). The two component problem is simulated by spectra in which peak A increases linearly with the concentration of component A and the maximum peak amplitude of B decreases quadratically with the concentration of component B. The simulated data set contains a training set of 36 spectra and a test set of 15 spectra. Each spectrum consist 500 spectral variables. In the design, the sum of the concentrations of component A and B was always equal to 1. So, the concentration and also the peak intensity of the components A and B are negatively correlated. Peak C is a randomly generated peak which is not related at all to any of the two other component concentrations. The same data set will also be used in Section 2.2.2 to demonstrate the SVR model interpretation method.

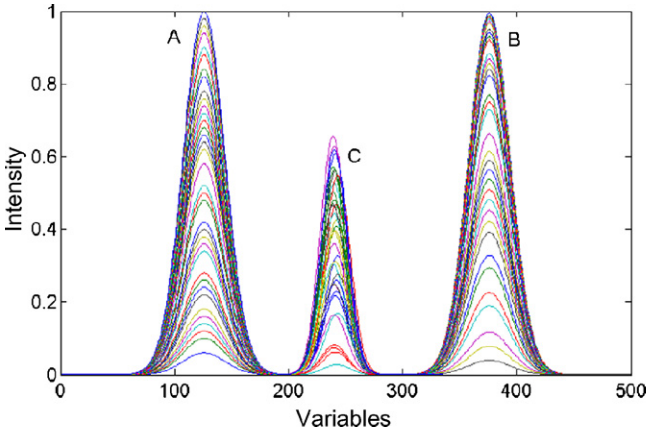

Fig. 3. The simulated spectra for the two component problem.

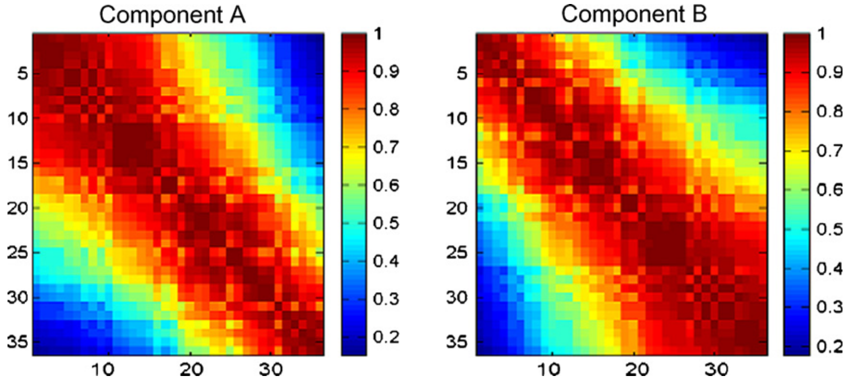

Fig. 4. Images of the two kernel matrices for the component A (left) and B (right) obtained by SVR using the PUK function. The concentration of component A and B increases going from left to right and up to down. The colour bar represents the Euclidean distance between the different objects weighted by the PUK function. A kernel value equal to 1 indicates a similar pair of objects, while a distance of 0.2 implies that a pair of objects is very dissimilar.

After determining the optimal SVR kernel its parameter settings by using the GA/Simplex optimisation procedure [9]. The training objects which are defined as non-support vectors (zero α-values) are eliminated before re-building the kernel matrix. The non-support vector objects have no contribution to the final model and therefore will disturb the visualisation of the kernel content. In case of the simulated data set all objects were defined as being support vectors. Fig. 4 depicts the images of the re-build kernel matrices for both components (A and B) of the simulated data. Since, SVR requires a separate model for each output property (Y-vector). Each model delivers its own kernel matrix and requires its own visualisation and interpretation. For this reason two models and thus two kernel matrices are obtained for the simulated data (compounds A and B). The objects of the kernel images in Fig. 4 are sorted by concentration per component and each row shows the weighted similarity between a specific object with that of the other objects. In both cases, objects corresponding to low concentrations exhibit dissimilarities with objects representing high concentrations (and the other way around).

Subsequently, the correlation matrices are constructed by calculating the standard correlation coefficient between the reduced input data set (hence, after elimination of the non-contributing vectors) and the re-build kernel matrices of both components as mentioned above. The correlation images together with the original spectra are depicted in Fig. 5. It can be seen that the regions located at the peak positions A and B give high correlation values (positive as well negative) for both components. While the region under peak C (the random peak) gives correlation values around zero. A high positive or negative correlation value suggests that these region(s) are relevant for the kernel construction, this in contrast to correlation values close to zero. In case of the simulated data, the sorted kernel matrices are mainly based on the peak regions of peaks A and B (indicated by the high correlation values), as theoretically could be expected.

The objects in the CI's are sorted by concentration and show the transition between the concentrations of both components A and B: if the concentration of component A increases the concentration of component B decreases. This effect is visualised in Fig. 5 by the oppositely correlation values between peaks A and B for both CI's. Furthermore, it is clear to see that the correlation value switches from a negative into a positive value, in case of component A. The same, but oppositely happens for component B. This effect can be explained by the sum of concentrations of components A and B, which is equal to 1. Finally, the horizontal light coloured bar in both CI's can be ascribed to the peak changes for the components A and B. The intensity of peak A increases, while the intensity of peak B decreases by the concentration. Since the total concentration must be 1, there will be a certain instance that both peaks have almost the same intensity which can be indicated as a kind of transition phase.

2.2.2. Interpretation of SVR models

In Section 2.2.1, we showed that it is possible to visualise the information which is embedded in the kernel matrix. It becomes clear which input variables in the original input data are explanatory for the modelled output property. To determine the contribution of each input variable to the final regression model, it will be useful to study the relation between the obtained α-values and the input variables. As mentioned before, the kernel matrix contains no direct information regarding the input vari-

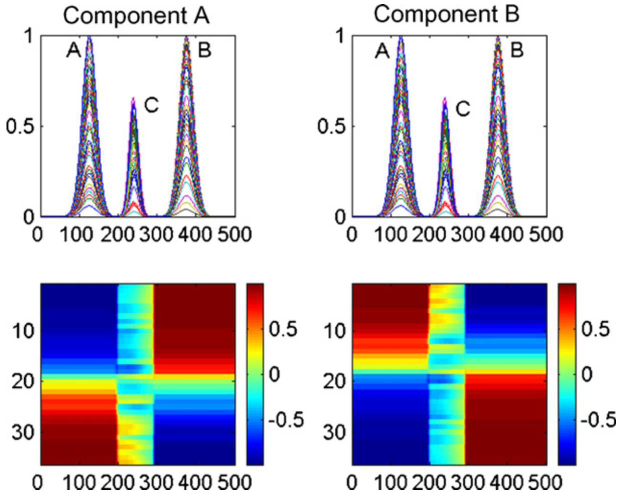

Fig. 5. The upper two plots depict the original spectra of the simulated two component problem. The lower part depicts the Correlation Images (CI) for both components, whereby the y-axis represents the objects sorted by the concentration (low-to-high concentration). The x-axis represents the input variables.

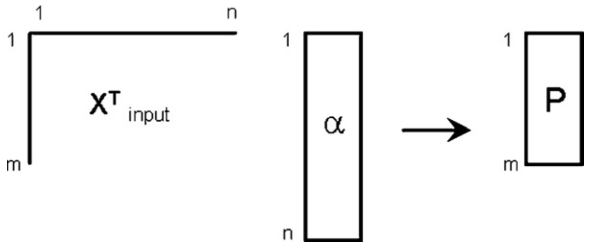

Fig. 6. A P-vector is obtained by calculating the inner-product between the original input space (X^T) and the α-vector of the SVR model.

ables; hence it makes no sense to relate the α-values to the kernel matrix rows. For this reason, the relation between the $\boldsymbol{\alpha}$-vector and the original input data is investigated.

The QP part of SVR returns a vector of α-values (which are comparable to the regression coefficients in the PLS algorithm) having a length equal to the number of objects (see Fig. 1), of which the elements satisfy the constraint $0 \leq \alpha_i, \alpha_i^* \leq C$, as was mentioned in Section 2.1. As a consequence each object of the original input space is weighted by its assigned α-value. The objects having α-values equal to zero are not important because these objects do not contribute to the SVR model, and, hence, must be eliminated before going further with the interpretation step. The reduced original input data set (thus containing the support vectors) and the corresponding non-zero α-values will be used to interpret the optimised SVR model. This is realised by calculating the inner product of the contributing input objects (reduced input data matrix $\mathbf{X}$) and the $\boldsymbol{\alpha}$-vector, which result a new vector ($\mathbf{p}$) with the length of the number of input variables, see Fig. 6. This vector yields a characteristic profile of the variables which contribute to the overall model.

Plotting of the obtained $\mathbf{p}$-vector in combination of the original input data makes it able to interpret the SVR model. The upper panel of Fig. 7 depicts the original input data. The lower row shows the $\mathbf{p}$-vectors corresponding to the models for the components A and B. Fig. 7 illustrates that the SVR model for component A is mainly based on the presence of peak A from the original input data and peak B is responsible for the SVR model of component B. This is exactly what we would expect considering the pre-defined structure of the data set. The peak intensity in both lower plots represents the contribution of each peak. The reason that peak B shows a small contribution for component A and peak A for component B depends on the fact that the total concentration of both components must be 1. Fig. 7 shows also the oppositely relation between components A and B. Namely, peaks A and B are negatively correlated.

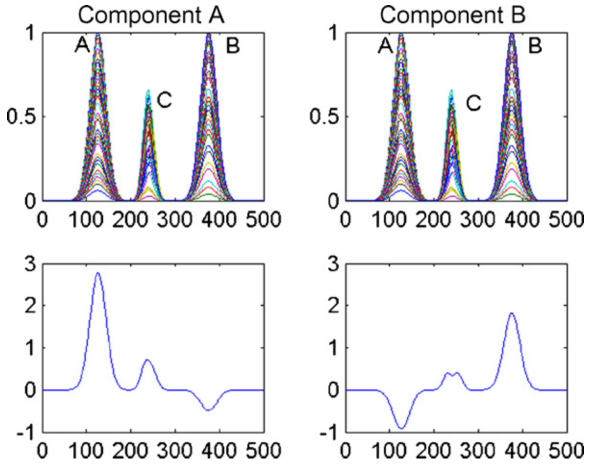

Fig. 7. The upper two plots depict the original spectra of the simulated two component problem. The lower part depicts the line plot of the $\mathbf{p}$-vector for component A and B.

It should be noted that still a weak reflectance is present for peak C in the P profile. This is due to the fact that only positive values are present in the simulated spectra: hence, the average profile of peak C contributes to some extent to the sum of the weighted input objects.

On the basis of those results it can be conclude that our feature space visualisation and SVR interpretation approaches are able to show which input variables of the original input space are responsible for the final SVR model. Especially, the interpretation of the SVR results, which corresponds with a loading plot of PLS, is a point of view that makes SVR understandable and will stimulate the use of SVR on industrial platform.

3. Data description

The two aspects, visualisation and interpretation of SVR, are investigated in this paper on the basis of three real-world data sets, which are described in detail in [9–11]. Furthermore, the SVR results and also the optimal parameter settings are already available from previous publications [9–11].

3.1. Metabolite data

The Metabolite data set, which is collected during the EC INTERPRET project [12] consists of 299 Nuclear Magnetic Resonance (NMR) spectra taken from well-specified locations in the brain of healthy volunteers as well as patients suffering from the tumour types oligodendroglioma (grade II (moderate disease) and grade III (server stage)). The spectra were taken over a range of 5.1–0.715 ppm resulting in 274 variables per spectrum and were corrected for a diversity of NMR artifacts like e.g., eddy currents [12]. For each spectrum the area under some pre-defined peak areas (corresponding to metabolites which are used by clinicians for the classification and grading of tumours) was calculated, resulting in six concentrations of the metabolites myo-inositol, choline, creatine, *N*-acetyl aspartate, lactate and fatty acids. Fig. 8 depicts the peak positions in the NMR spectra which correspond to the mentioned six metabolites.

A training set of 197 NMR spectra and a test set of 102 NMR spectra are selected from the original data set (299 NMR spectra) to predict the accompanying metabolite concentrations by SVR. It should be stressed here, that this is a typical linear regression problem because each metabolite concentration was derived by a straight numerical integration over the associated peak area.

3.2. Methanol distillation data

The methanol distillation data set consists of a total of 131 NIR spectra measured in a temperature range of 20–40 °C and

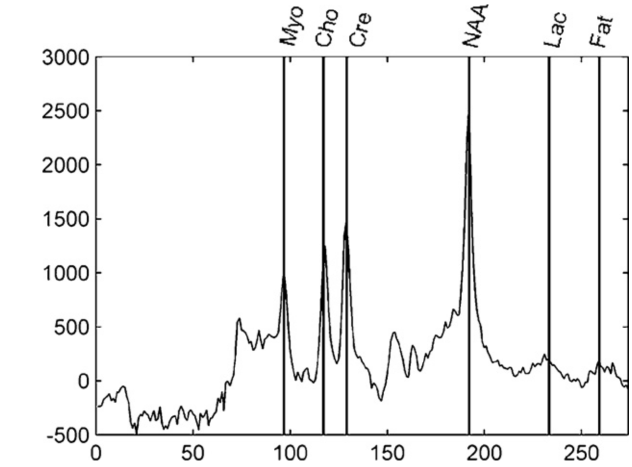

Fig. 8. Typical NMR spectrum for healthy brain tissue. Vertical bars indicate the peak positions corresponding to myo-inositol (Myo), choline (Cho), creatine (Cre), *N*-acetyl aspartate (NAA), lactate (Lac) and fatty acids (Fat), respectively. Note the absence of lactate and fatty acids peaks in the spectrum: these are not present in healthy tissue.

in a spectral range of 4000–10,000 cm^{-1}. A large part of these methanol samples is measured under controlled laboratory circumstances (89), while the remaining part is directly collected *in situ* from the plant (46). As described in the original paper [9], the laboratory samples are used as training set to predict the water impurity in the plant samples (the independent test set). The SVR settings and results published in [6] are used as a benchmark.

3.3. Corn data

The corn data set consists of 80 NIR spectra of corn samples to estimate the amount of moisture, oil, protein and starch contents. The spectra were measured from 1100 to 2498 nm at a spectral resolution of 2 nm on three different NIR spectrometers (indicated by the acronyms: m5, mp5 and mp6). A more detailed description including the data set itself is available on the website of Cargill (http://software.eigenvector.com/Data/Corn/index.html). In this paper, the NIR spectra measured on the 'm5' spectrometer and the SVR settings which are published in [6] are used to visualise and interpret the constructed SVR models. The same data pretreatment and training and test set selection has been performed as was described in the aforementioned reference.

3.4. Software and generalisation criterion

The SVR calculations were performed using the SVM toolbox developed by Gunn [13], which can be downloaded from http://www.isis.ecs.ac.uk/isystems/kernel/. Optimisation of the SVR parameters (kernel parameters, ε range and the regularisation constant C) are based on the Genetic Algorithm/Simplex optimisation approach as has been outlined in [9]. All programs are implemented in Matlab V6.5 (The Mathworks Inc.) and carried out on an Intel Pentium IV 3.0 GHz with 1 GB of memory and the Windows XP operating system.

The root mean square error of prediction (RMSEP) is used as a performance criterion calculated from the measured and predicted values of the SVR model. The RMSEP is defined by:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \qquad (4)$$

where $\hat{y}_i$ and y_i denote the predicted value and the measured value, respectively. Here, n denotes the number of objects in the test set.

4. Results and discussion

4.1. Metabolite data

The visualisation and interpretation approach described in Section 2.2 will be used here to show its power and possibilities for analyzing the Metabolite data set. On beforehand, it is known that the myo-inositol, choline, creatine and the *N*-acetyl aspartate metabolites are weakly correlated but not to the lactate and fatty acid concentrations. However, the lactate and fatty acids are correlated to each other, because a high-grade tumour contains lactate as well as fatty acids. Assuming that there is only a weak correlation between the metabolites, we expect that each peak in the NMR spectra, excepting the lactate and fatty acids peaks, will be more or less uniquely responsible for the associated metabolite concentration. For example, the myo-inositol concentration will only be affected by the Myo peak in the NMR spectra (Fig. 9). Since SVR builds a model for each metabolite concentration separately, a unique model, we must be able to see this effect by applying the SVR interpretation procedure that is described in Section 2.2.

Table 1 summarises the optimal parameter settings and the performance of the SVR model for each metabolite concentration. Since the distance based PUK kernel function is used. It is possible to make a statement if the obtained models are linear or non-linear, based on the σ-value of PUK and the maximal distance value between the objects. For the metabolite data, the maximal distance between the objects of the training set is 1.29×10^4. The ratio distance/σ-values are in the magnitude of 700, which indicates that the obtained models are linear. From Table 1 can be seen that in all cases SVR is able to generate good results (a low RMSEP value together with a high correlation coefficient (R)).

The following step now is to understand the driving force underlying these results. For doing this, the obtained SVR models are analysed by the visualisation and interpretation procedure described in Section 2.2.

The upper part of Fig. 9 depicts the spectra of the training set, which is used the build the SVR models. The other panels display the correlation image for each metabolite, which is obtained by calculating the correlation between the training set and the kernel matrix (characteristic to each metabolite) generated by using the kernel settings given in Table 1. As can be observed, the six correlation images are almost similar in the sense that the same spectral regions are defined as most relevant (regions with the highest correlation values (positive as

Table 1
SVR parameter settings and SVR prediction results of the metabolite data set

	Myo	Cho	Cre	NAA	Lac	Fat
σ	17.1	18.7	16.7	17.1	21.9	20.6
ω	79	42	28	3	17	44
ε	0.2	0.0162	0.2	0.1039	0.0923	0.1552
C	7.74×10^6	1.13×10^4	3.43×10^4	9.87×10^3	2.36×10^4	2.69×10^4
RMSEP	33.1	37.9	34.3	48.7	22.6	33.6
R	0.9951	0.9964	0.9926	0.9949	0.9981	0.9898

well negative)). However, the peak that is really related with the concerning metabolite shows an explicit gradient in the colour intensity in the correlation images at that corresponding peak position. For example, the correlation image of lactate shows a nice gradient in colour intensity for the variables between 230 and 250, which is not clearly visible in the other five correlation images (at same spectral range). This can be explained by the fact that the objects in the correlation images are sorted by each metabolite concentration. Thus, the peak(s) that are highly correlated with the concerning metabolite concentration will show a clear gradient. However, this means not that peak(s) without a certain gradient are not always meaningful. From the correlation images of Fig. 9, we can deduce that the six spectral peaks related to the six metabolite concentrations are manifest as relevant. Those peaks show a clear gradient and are displayed by an intense colouring (high correlation value). The regions that are non-explanatory are displayed by a light colouring (correlation values close to zero).

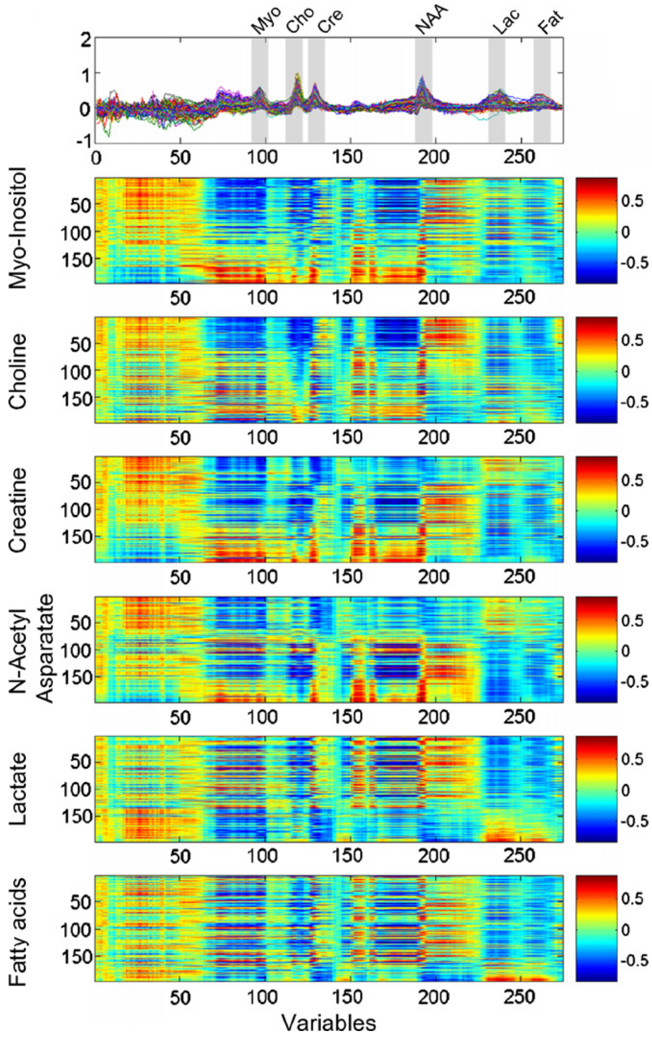

Fig. 9. The upper panel depicts the original input spectra of the metabolite data. The vertical grey bars indicate the numerical integration regions of the six metabolites concentrations in the order myo-inositol, choline, creatine, *N*-acetyl aspartate, lactate and fatty acids. The other panels depict the correlation images between the original input spectra and the kernel matrix of each SVR model for the six metabolites concentrations. The rows of each correlation image are sorted by the respective metabolite concentration. The *x*-axis is given in arbitrary units.

Summarizing, by applying the kernel formalism, SVR appears to be capable to discriminate between regions that are meaningful or non-explanatory for the associated output property (e.g., each individual metabolite concentration).

Fig. 10 depicts the NMR spectra of the training set and the line plots of the P-vector of each SVR model, which is obtained by calculating the inner-product between the SVR **α**-vector and the training set as was described in Section 2.2. It can be seen that each SVR model concentrates on a specific spectral region. A closer examination of Fig. 10 shows that that these regions correspond to the on beforehand defined integration regions depicted in Fig. 8. Hence, we can conclude that the SVR model indeed extracts the most relevant spectral features for modeling the respective metabolite concentration. Fig. 10 also confirms our assumption that each metabolite concentrations will be affected by certain spectral regions as defined in Fig. 8. However, we could expect that lactate and fatty acids should be correlated [11]. This means that the line plots of the lactate and fatty acids SVR models should be based on the lactate as well as fatty acids peaks, but this is not the case. However, the correlation images shown in Fig. 9 illustrate that the lactate and fatty acids concentrations are correlated. In both cases the lactate as well as the fatty acids peak regions show a nice gradient in colouring. Thus, both peaks are related to the lactate as well as fatty acids concentration. The reason why this is not visible in the line plots can be explained by the fact that SVR is capable to build a model which diversifies exactly to the specific metabolite concentration.

Summarizing, visualisation of the feature space depicts the relevant spectral regions, while the interpretation depicts the spectral regions which are directly correlated to the concerning output property. The obtained results confirm that our visualisation and interpretation approach is facilitating to understand the modeling power of SVR. In other words, a SVR model is not a black box anymore.

4.2. Methanol distillation data

In two previous publications [6,9], we showed that SVR is capable to model the methanol distillation data to predict the percentage of water impurity in methanol by using NIR spectra. The optimal SVR parameter settings and the prediction results from literature [6] are recapitulated in Table 2. Here, again we can make the statement if the SVR model is linear or non-linear. The maximal distance between the training objects is 11.7. A ratio of 0.6 will be obtained for distance/σ, which indicates that we are dealing with a non-linear model. Examination of this table leads to the conclusion that SVR results in a prediction model with a high problem-solving power (a correlation coefficient (*0*) almost equal to 1). However, at that time we were not able to explain why SVR was so successful. We did use SVR more or less like a black box and were only interested in its performance. Now we are able to interpret the SVR model and to explain why SVR was so effective for this problem.

Table 2
SVR parameter settings and prediction results of the methanol data set

	Water
σ	19.5
ω	0.4
ε	0
C	1.48×10^4
RMSEP	0.0222
R	0.9998

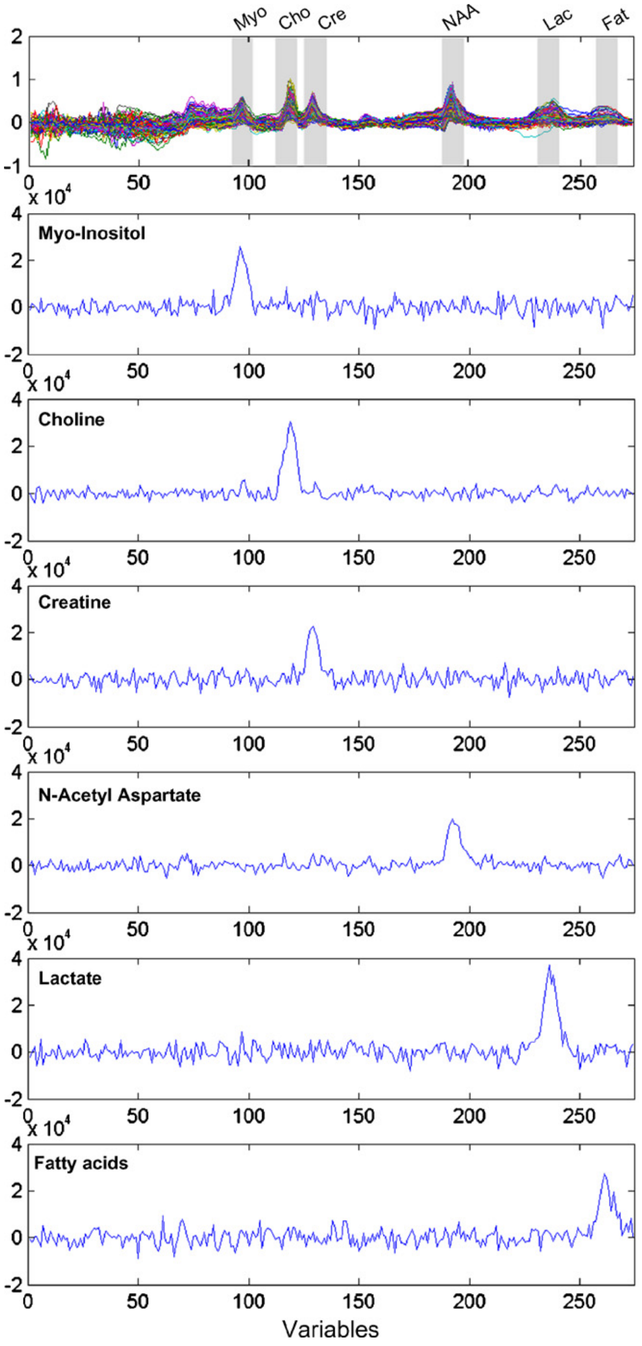

Fig. 10. The upper panel depicts the spectra of the training set. The other panels depict the inner-product between the spectra of the training set and the **α**-vector of each SVR model. The *x*-axis is given in arbitrary units.

The upper part of Fig. 11 depicts the spectra that are used as training set. A visual observation shows that the spectral region between 0 and 500 together with 1000–1500 can be related to the water impurity level. To know if this observation is correct and to examine if the SVR model is really based on these spectral regions we have analysed the SVR model by our visualisation and interpretation approach. The lower panel of Fig. 11 shows the correlation image. As can be seen, both mentioned spectral regions are defined as being relevant for the SVR model (indicated by high correlation values; positive as well negative). Furthermore, the spectral regions around 780, 900 and 2000–3000 appear to be meaningless for the SVR model (the bands, representing low correlation values). This might be caused by the fact that the spectral regions between 780 and 900 contain some non-linear temperature induced spectral shifts. Apparently, the presence of these spectral shifts has been ruled

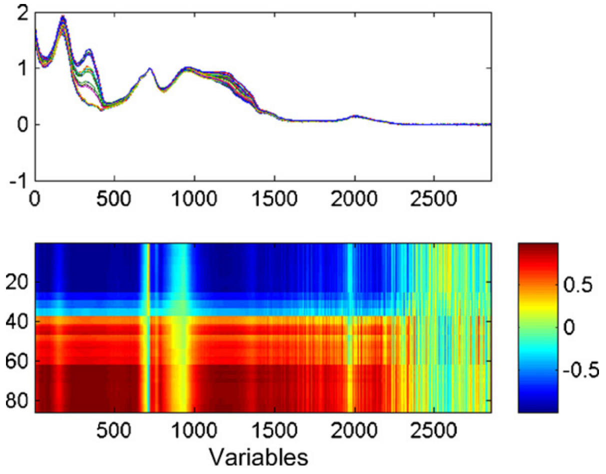

Fig. 11. The upper panel depicts the spectra of the training set. The lower panel depicts the correlation image. The objects are sorted by concentration (*y*-axis). The *x*-axis is given by arbitrary units.

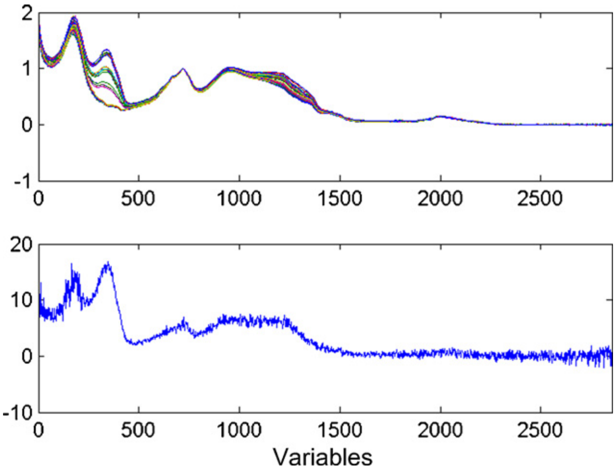

Fig. 12. The upper panel depicts the spectra of the training set. The lower panel depicts the inner-product between the training set and the α-vector of the SVR model. The *x*-axis is given by arbitrary units.

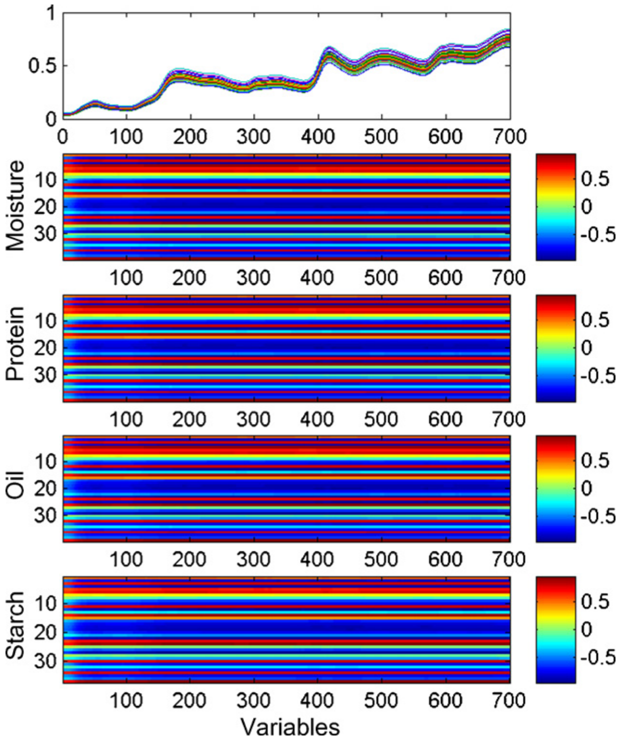

Fig. 13. The upper panel depicts the original input spectra. The other panels depict the correlation images obtained by the original input spectra and the kernel matrices of the moisture, protein, oil and starch components. The matrix rows are sorted for each image by the concentration of the considered component. The *x*-axes are given in arbitrary units (variable numbers).

out by the kernel transformation. Moreover, the spectral range between 2000 and 3000 seems to contain no relevant information at all, which can be attributed to the small spectral deviations in that region in relation the variation in the water concentration.

Indeed, by interpreting the SVR model it is clear that the spectral region >1500 has no contribution to the concentration of the water impurity, see Fig. 12. This is more or less what we could expect on basis of the correlation image. It can be concluded that the SVR model is exclusively based on the information that is available in the spectral region between 0 and 1500 Moreover the spectral region 0–500 shows a higher intensity in comparison to the spectral region between 500 and 1500. Probably this region will also have the highest contribution to the SVR model. Namely, theoretically water will give always a broad resonance band in the region 0–500 for NIR spectra. Probably, the most intense information about water is present in this region. However, confounding effects may play a role as well. Hence, further research is needed to confirm this observation.

From the above results it can be concluded that the prediction results of SVR are based on the spectral information that is located in the region 0–1500. Moreover, probably it is possible to make a distinction between the contributions of the different spectral regions. This can be used to apply feature selection to improve the prediction performance, to simplify the model and, possibly, make the model more robust.

Regarding our earlier publications [6,9], we can now interpret the optimised SVR model and are able to understand why SVR is successful in solving this problem in an adequate way.

4.3. Corn data

The corn data is a typical example of a spectral data set where it is not possible to make on beforehand decisions which spectral region or regions are informative to explain the four components (concentrations of moisture, oil, protein and starch, respectively). The intensity of the whole spectrum increases or decreases by an increase or decrease of the concentration of each component, see upper panel Fig. 13. Thus, visually it is not possible to assign specific spectral regions to each of the four components. However, SVR is able to predict the concentrations of these components very good (Table 3). Hence, the input data must contain specific information regarding the concentration of these components, which should be become clear by our SVR interpretation approach.

The maximal distance between the training objects is 3.5. The distance/σ ratios lie between 0.03 and 0.09 for the different SVR models. This indicates that we are dealing with non-linear models.

The lower panels of Fig. 13 depict the correlation image of each SVR model of moisture, oil, protein and starch contents. Visual inspection of the correlation images indicates that all variables per object have the same contribution (same colouring) to the kernel matrix. However, a numerical investigation makes it clear that there are some small differences, which are not big enough to make discrimination between the spectral regions.

Table 3
SVR parameter settings and prediction results of the Corn data set

	Moisture	Protein	Oil	Starch
σ	40.4	155.6	129.9	130.4
ω	100	71	171	25
ε	0	0.001	0	0.064
C	1.6×10^7	1.0×10^8	5.3×10^7	1.0×10^8
RMSEP	0.0099	0.1132	0.0637	0.1665
R	0.9998	0.9702	0.9586	0.9784

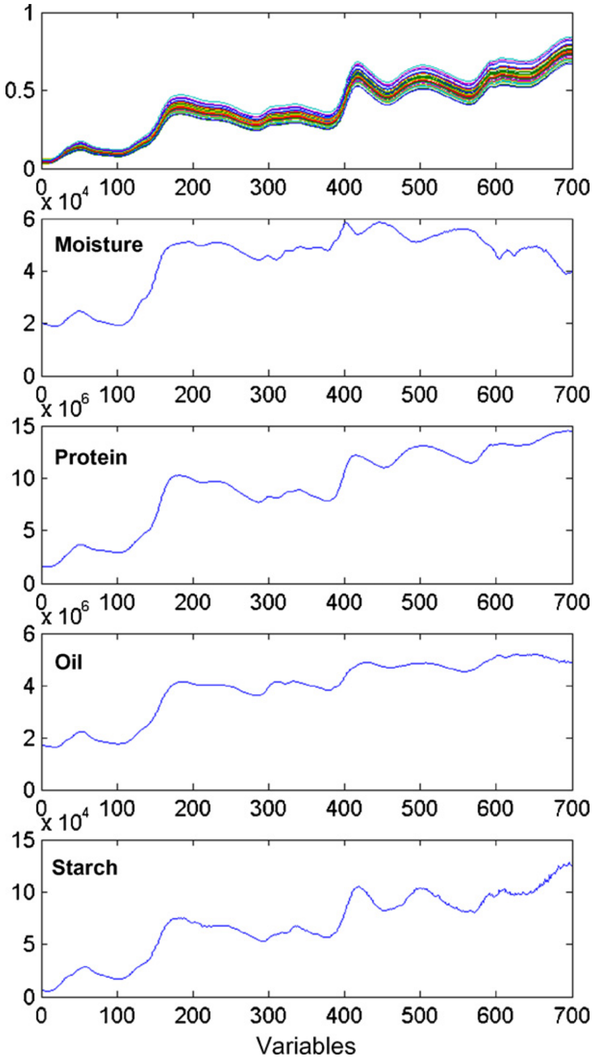

Fig. 14. The upper panel depicts the original input spectra. The other panels show the inner-product profiles obtained between the original input spectra and the **α**-vector of the different SVR models in the order moisture, protein, oil and starch, respectively. The *x*-axes are given in arbitrary units.

Therefore, as expected, the whole spectral region is indicated as being explanatory to drive the predictive model. Furthermore, the correlation images are more or less similar and do not show any gradient in colour intensity for the four contents. This might be caused by the fact that those four contents are strongly correlated. A corn sample will contain moisture, protein, oil as well as starch together. So, the NIR spectra contain just some ‘lumped sum’ of the overall information: sorting of the rows of the CI matrix by concentration per component will not result in a gradient like colour profile. Most of the objects of the four correlation images are intensely coloured, which suggests that these objects are explanatory. This information can be used to eliminate some of the objects in the training set. Again, further research is necessary to confirm this observation.

Interpretation of the SVR models (Fig. 14) shows us that there are some spectral differences for the four contents that are responsible for the behaviour of the SVR model. The main profile is almost the same for the four components. However, there are some small spectral differences between 270 and 700. Especially, the spectral region between 580 and 700 is different for the moisture, oil and protein, starch components. In case of moisture and oil the spectral amplitude decreases in intensity, while in case of protein and starch the intensity increases. Those differences possibly contribute to the predictive power of the SVR model.

In conclusion, SVR is able to extract powerful information from this spectral data set. The proposed interpretation approach described in this paper makes it possible to visualise this which was not visible by eye on beforehand.

5. Conclusion

In this paper, we presented a visualisation and interpretation tool to turn SVR, which is mainly used as a black box, into a transparent and interpretable modeling technique. Our analysis tool has two main advantages: first, the visualisation of the feature space shows which variables in the original input data are related to the associated output variable. In other words, which part(s) of the input data includes information which can be useful to estimate the accompanying output value? Second, the interpretation step of the final SVR model makes it clear which input variables are explanatory for the good performance of SVR. Both steps (visualisation and interpretation) together helps to make the SVR model transparent to the user and this will hopefully stimulate the use of SVR for tackling new non-linear regression problems.

The efficiency and effectiveness of our visualisation and interpretation approach is illustrated through performing this analysis on a simulated and three real-world data sets. The results showed that our approach is capable to visualise the content of the kernel matrix to understand the transformation of the original input space into the feature space. It visualises the information derived from the original input space. Our analysis makes it possible to decide which input variables have a high influence on the kernel matrix and are responsible for the predictive power of SVR. The visualisation of the SVR kernel matrix, and especially the interpretation of the optimised SVR model can be used as a feature selection technique. The visualisation and interpretation steps indicate which input variables are on the one hand correlated and on the other hand uncorrelated to the output variable. This information can be used to eliminate some parts, i.e., redundant or non-explanatory variables, from the original input data set. Furthermore, the intensity of the peaks present in the **p**-profile can maybe helpful to make a distinction between the contributions of the different spectral regions. These issues will be the subject of a forthcoming paper.

In the past years, an extensive interest has been shown in developing and applying kernel-based methods which are suitable for solving non-linear regression and classification problems. An example is Kernel–Partial Least Squares (K-PLS) [14–16]. PLS by itself is known as being a transparent modeling technique which has the possibility to interpret the final regression model by its scores and loadings. The PLS loadings indicate which input variables are strongly correlated to the predicted output value(s). However, by using the kernel formalism this interpretation possibility is lost for the same reason as was discussed for SVR. Our visualisation and interpretation approach, which intentionally was developed for SVR, can also be applied to interpret a K-PLS model. For example by replacing the **α**-vector of the SVR mode by the **β**-regression coefficient vector of the K-PLS model [16]. The application of the visualisation and interpretation approach on K-PLS and also on kernel based classification techniques e.g., Support Vector Classification (SVC) will be investigate.

In summary, it can be concluded that the proposed visualisation and interpretation approach described in this paper facilitates to understand and analyse the build SVR models in depth. To our opinion, this will boost the application of SVR models to tackle a broad gamut of non-linear regression problems.

Acknowledgement

This work has been funded by the EU-project TRACE (No. 006942, http://www.trace.eu.org).

References

[1] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, 1998.

[2] V. Vapnik, The Nature of statistical Learning Theory, Springer-Verlag, New York, USA, 1995.

[3] V. Vapnik, Statistical Learning Theory, John Willey & Sons, New York, USA, 1998.

[4] B. Schölkopf, A.J. Smola, Learning with Kernels, MIT press, Cambridge, 2002.

[5] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, Cambridge, UK, 2000.

[6] B. Üstün, W.J. Melssen, L.M.C. Buydens, Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel, Chemom. Intell. Lab. Syst. 81 (2006) 29–40.

[7] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore, 1999.

[8] R. Fletcher, Optimization, London, Academic Press Inc., 1969.

[9] B. Üstün, W.J. Melssen, M. Oudenhuijzen, L.M.C. Buydens, Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization, Anal. Chim. Acta 544 (2005) 292–305.

[10] R.N. Fuedale, H. Tan, D. Brown, Piecewise orthogonal signal correction, Chemom. Intell. Lab. Syst. 63 (2002) 129–138.

[11] W.J. Melssen, B. Üstün, L.M.C. Buydens, SOMPLS: a supervised Self-Organising Map—Partial Least Squares algorithm for multivariate regression problems, Chemom. Intell. Lab. Syst. 86 (2007) 102–120.

[12] A.W. Simonetti, W.J. Melssen, F. Szabo de Edelenyi, J.J.A. van Asten, A. van Heerschap, L.M.C. Buydens, Combination of feature-reduced MR Spectroscopic and MR imaging data for improved brain tumor classification, NMR Biomed. 18 (2005) 34–43.

[13] S.R. Gunn, Support Vector Machines for classification and regression, in Technical Report, Image Speech and Intelligent Systems Research Group, 1997, University of Southampton: Southampton, UK.

[14] B. Walczack, D.L. Massart, Application of Radial Basis Functions—Partial Least Squares to non-linear pattern recognition problems: diagnosis of process faults, Anal. Chim. Acta 331 (1996) 177–185.

[15] B. Walczack, D.L. Massart, The Radial Basis Functions—Partial Least Squares approach as a flexible non-linear regression technique, Anal. Chim. Acta 331 (1996) 177–185.

[16] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in reproducing Kernel Hilbert space, J. Machine Learn. Res. 2 (2001) 97–123.