# Title page

Comparison of multivariate regression techniques and variable selection methods for near infrared spectroscopy of articular cartilage

Prakash M1, Sarin J.K1,2, Rieppo L1,3, Afara I.O1,2,4, Töyräs J1,2

1 Department of Applied Physics, University of Eastern Finland, Kuopio, Finland

2 Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland

3 Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland

4 Department of Electrical and Computer Engineering, Elizade University, Ondo, Nigeria

**An abbreviated title:** Comparison of regression methods in Equine AC NIR data

**Planning to submit to:** Analytica Chimica [Acta](http://www.journals.elsevier.com/analytica-chimica-acta) September 2016

**Corresponding author:**

Mithilesh Prakash, M.Sc. (Tech.)

Department of Applied Physics

University of Eastern Finland

Kuopio, Finland

Tel: +358 414963056

Fax: +35817162131

Email: [mithilesh.prakash@uef.fi](mailto:mithilesh.prakash@uef.fi)

# Abstract

Near infrared (NIR) spectroscopy has been successfully applied to non-destructive assessment of the properties of biological tissues, such as articular cartilage and proposed to be used during clinical arthroscopies. In general, the spectroscopic data includes absorbance values from a broad wavelength region, thus resulting in a large number of contributing factors and complexity amongst variables. Furthermore, this broad spectrum includes information from noisy variables potentially contributing to errors during spectroscopic analysis. We hypothesized the partial least squares regression (PLSR) to be optimal regression method and variable selection methods to significantly increase the reliability of the results. To test the hypothesis in this study, for the first time a comparative analysis of multivariate regression methods in NIR spectroscopy on cartilage data is presented and the effect of variable selection methods on the performance of PLSR models evaluated. The comparative analysis indicates PLSR technique to be an ideal choice for cartilage NIR research. Furthermore, variable selection methods, especially monte carlo uninformative variable elimination(MC-UVE) with PLSR enhanced the results. Thus, variable selection based PLS can effectively model cartilage NIR data with functional tissue parameters.

**Key terms:** spectroscopy; variable selection; near infrared; partial least squares (PLS) regression

# Introduction

Articular cartilage (AC) is a specialized type of hyaline cartilage found at the distal ends of bones providing smooth low friction load bearing interfaces in joints. Degenerative joint conditions such as Osteoarthritis (OA) are generally characterized by the disruption of the superficial collagen network, the loss of proteoglycans (PGs) and changes in tissue functional properties[1–3]. Non-destructive light-based spectroscopic imaging modalities, such as Optical coherence tomography (OCT)[4], Fourier transform infrared (FTIR)[5], Near infrared (NIR) and Raman[6], have been quite effective in cartilage research. Recent cartilage research studies ( Afara et al [4,7], Palukuru et al [8] and McGoverin et al [9]) have advocated the use of NIR as a better method for in vivo studies due to its superior tissue depth penetration. Application of NIR spectroscopy in AC research has enabled the assessment of the tissue compositional and biomechanical properties in OA [7,10,11].

Multivariate[12] regression techniques are typically used to extract information from the NIR spectral data since univariate approaches are often inefficient due to multi-collinearity in NIR data. Regression methods enable building mathematical models relating NIR spectra with the functional properties of the sample specimen (e.g. cartilage). The popular multivariate regression techniques utilized in NIR spectroscopy are principal component regression (PCR)[13] and partial least squares regression (PLSR)[14]. PLSR is most commonly used in NIR spectroscopic studies of articular cartilage. However, the regression shrinkage methods, such as ridge regression[15] and least absolute shrinkage and selection operator (LASSO), and least square version of support vector machines based regression called LS-SVM[16], have not been used in NIR studies of articular cartilage.

Selecting optimal variables for regression models is an essential step as the spectra may contain noisy or irrelevant variables that hinder the analysis. Usually variable selection has been done by restricting the spectral wavelength due to experimental or known restrictions (also called manual selection of wavelength range), which may result in inconsistent results and is prone to human error. Statistical studies conducted by Xiaobo et al[17], Westad et.al[18] and Mehmood et.al[19] have shown the significance of variable selection methods in multivariate regression techniques. In general, variable selection[19] in regression is based on the principle of either choosing the most contributing variables or eliminating the noncontributing variables[17]. MC-UVE, CARS, VCPA, interval selection methods, GA and jack-knife encompass different variable selection methods available for analyzing NIR spectrum.

In this study, multiple multivariate regression and variable selection methods are utilized in an attempt to find the most suitable algorithms for cartilage research. We hypothesized the PLSR to be optimal regression method and variable selection methods to significantly increase the reliability of the results. To test the hypothesis, calibration models between spectral and reference data were built and evaluated with an independent test group

# Materials and methods

1. *Material*

The study was conducted using NIR data and determined from equine cartilage in our earlier study. Briefly, equine metacarpophalangeal joints (*N=5*) were gathered and areas of interest (*AI, N=44*) of intact and damaged cartilage were selected by equine surgeons. The AIs were uniformly divided into grids, thus resulting in a total of 869 measurement points. The spectral data included the absorbance information from the wavelength range of 700-1050 nm.

The tissue properties including equilibrium, dynamic and instantaneous moduli were obtained via indentation[20] testing. Additionally, cartilage thickness was determined with optical coherence tomography[21]. The details of the data acquisition are given in *Sarin et al* [22].

1. *Methods*

The NIR spectra data was preprocessed by smoothing and filtering using a third degree Savitzky-Golay filter with a 25 nm window to remove the background noise. Furthermore, second derivative was applied on the smoothed and filtered data to remove the offset and linear elements of the baseline and to increase the separation between the absorption peaks.

The dataset was divided into two sets similarly as in [22]. The first set, called the training set, consisting of 799 samples was used for calibration model training, and the second set, called the test set, consisting of 70 samples was used to evaluate the model performance[23]. After the technique specific optimization, the model parameters were finally optimized for the highest *R2* and the lowest root mean square error of prediction (RMSEP) for the test set. The software analysis was done using MATLAB R2014a (8.3.0.532).

*Regression methods*

PCR, PLS, ridge, LASSO and LS-SVM methods were used for multivariate regression comparative analysis. Built-in MATLAB functions *princomp, plsregress, ridge, lasso* and *LS-SVM toolbox* from LS-SVMlab[24] v1.8 were used to regress the NIR data with functional parameters respectively.PCR (*princomp*) was optimized by first building series of models with iteratively increasing number of principal components (PC). The number of PC’s were varied by increasing the number principal component scores and principal component loadings simultaneously until maximum number of components was reached (max =15). Next, the best PCR model with the highest *R2* test and the lowest RMSEP was retained and the number of components recorded. Similarly, PLSR (*plsregress*) was first optimized by building series of models as function of number of components (*maximum components =15*) and each model cross-validated individually by *k*-fold operation (*10 fold*) and the best PLSR model with the highest *R2* test set and the lowest RMSEP retained. For Ridge regression, the vector of the ridge parameters (0 to 1000) was varied in small steps (step size = 0.01) and series of models built as a function of the step size. Similar to PCR and PLSR, the best model was retained. In LASSO regression, the initial coefficients are calculated by regularization algorithm (*lasso*) and cross-validated by k-fold (k=10). Models series were built as a function of variation in penalty term (βlasso). Following the approach used in PCR, PLSR and ridge regression, the best model was retained. For LS-SVM the initialization (*initlssvm*), tuning (*tunelssvm*) and optimization (*trainlssvm*) of the preprocessed is auto handled by the respective functions of the toolbox.

*Variable selection methods*

MC-UVE, CARS, VCPA, biPLS, GA and Jack-knife were used to further enhance the PLSR prediction models. The algorithms for MC-UVE and CARS[25] were obtained from *Integrated library for PLS and discriminant analysis*[26], VCPA[27] from *Variable Combination Population Analysis toolbox*, GA algorithm from *PLS-Genetic algorithm toolbox*[28] and Jack-knife algorithm[29] was coded in-house.

MC-UVE (*mcuvepls*) variable selection was optimized by first calculating the reliability index of all the wavelengths and then determining the optimal threshold for reliability index by finding the maximum correlation with the training set. CARS (*carspls*) and VCPA (*vcpa*) did not require additional input and the respective functions auto handled the optimization. In interval selection method the algorithm is optimized by elimination of the non-informative intervals and the 3 intervals (*intervals = 5, 17, 19*) with the lowest RMSECV were retained. In GA algorithm the effective number of evaluation n1 and number of variables was first evaluated n2, 182 and 100 respectively, using the *gaplsopt* function in the toolbox and the main function *gaplssp* was invoked to perform the GA algorithm. In Jack-knife the student’s t-statistics is used to determine the variable selection by selecting variables with values less than the predefined threshold (*t = 0.05*).

*Statistics and model comparison*

The calibration models developed were analyzed on five key parameters: root mean square error of calibration (RMSEC), R2 training set, root mean square error of prediction (RMSEP), R2 testing set and error percentage in the test dataset. Additionally, for PCR and PLSR models, the number of components are recorded.

# Results

The regression models were built and optimized for cartilage thickness, instantaneous modulus, equilibrium modulus and dynamic modulus. PLSR technique was found to the best with the highest R2 test and the lowest RMSEP and error percentage amongst the studied regression methods in case of cartilage thickness, instantaneous modulus and equilibrium modulus (table 1). In case of dynamic modulus property, LASSO was found to have the highest *R2* for the training set, but it also had a higher error percentage than PLSR.

**Table 1**: Comparison of different regression methods across different tissue parameters. Data is arranged in the descending order for the R2 Test set as highlighted in bold.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Cartilage Thickness (mm)** | | | |  |
| **Regression method** | **Train Set** | | **Test Set** | | **Error %** | |
| **R2 Train** | **RMSEC** | **R2 Test** | **RMSEP** |
| **PLSR (C\* = 5)** | 70.28 | 0.13 | **75.57** | 0.11 | 5.94 | |
| **RIDGE** | 73.02 | 0.12 | **74.09** | 0.11 | 6.17 | |
| **LASSO** | 72.90 | 0.12 | **68.63** | 0.12 | 6.90 | |
| **LSSVM** | 77.55 | 0.11 | **67.87** | 0.13 | 6.89 | |
| **PCR (C= 13)** | 60.44 | 0.15 | **67.38** | 0.13 | 7.02 | |
|  | **Instantaneous Modulus (MPa)** | | | |  |
| **PLSR (C = 6)** | 41.82 | 2.63 | **57.16** | 2.31 | 9.04 | |
| **RIDGE** | 41.82 | 2.63 | **54.08** | 2.39 | 9.76 | |
| **LASSO** | 40.82 | 2.65 | **52.67** | 2.42 | 9.79 | |
| **PCR (C = 10)** | 29.60 | 2.89 | **51.50** | 2.45 | 9.76 | |
| **LSSVM** | 98.97 | 0.35 | **43.78** | 2.64 | 11.41 | |
|  | **Equilibrium Modulus (MPa)** | | | |  |
| **PLSR (C = 5)** | 65.95 | 0.87 | **66.84** | 0.97 | 15.73 | |
| **LASSO** | 82.03 | 0.63 | **60.51** | 1.06 | 17.36 | |
| **RIDGE** | 75.88 | 0.73 | **54.54** | 1.17 | 20.39 | |
| **LSSVM** | 60.28 | 0.93 | **49.27** | 1.20 | 20.66 | |
| **PCR (C = 15)** | 24.74 | 1.29 | **35.18** | 1.36 | 21.63 | |
|  | **Dynamic Modulus (MPa)** | | | |  |
| **LASSO** | 69.17 | 3.44 | **66.35** | 3.56 | 13.85 | |
| **PLSR (C = 2)** | 37.27 | 4.91 | **64.88** | 3.64 | 13.06 | |
| **RIDGE** | 63.46 | 3.75 | **61.31** | 3.83 | 15.44 | |
| **PCR (C = 6)** | 27.03 | 5.30 | **60.35** | 3.87 | 14.55 | |
| **LSSVM** | 72.83 | 3.23 | **59.08** | 3.93 | 15.43 | |

*\*Number of components for PLSR and PCR is indicated by C*

**Table 2**: Comparison of variable selection methods. The data is presented in the descending order of *R2* for the test set as highlighted in bold.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Cartilage Thickness (mm)** | | | | |  |
| **Variable Selection method** | **No. Of PLS Components** | **Train set** | | **Test set** | | **Error Percentage** | |
| **R2** | **RMSEC** | **R2** | **RMSEP** |
| **MC-UVE-PLS** | 4 | 70.53 | 0.14 | **76.20** | 0.11 | 5.89 | |
| **JK-PLS** | 1 | 59.65 | 0.15 | **74.05** | 0.11 | 6.14 | |
| **GA-PLS** | 6 | 70.65 | 0.15 | **70.76** | 0.12 | 6.47 | |
| **BIPLS** | 12 | 59.05 | 0.15 | **70.05** | 0.12 | 6.59 | |
| **VCPA** | 3 | 63.15 | 0.15 | **66.13** | 0.13 | 7.02 | |
| **CARS** | 4 | 70.58 | 0.14 | **62.02** | 0.14 | 7.65 | |
|  | **Instantaneous Modulus (MPa)** | | | | |  |
| **CARS** | 5 | 43.39 | 2.79 | **56.72** | 2.32 | 8.99 | |
| **MC-UVE-PLS** | 4 | 39.19 | 2.86 | **56.66** | 2.32 | 9.46 | |
| **JK-PLS** | 7 | 32.40 | 2.98 | **53.21** | 2.41 | 9.41 | |
| **GA-PLS** | 2 | 30.97 | 2.88 | **51.56** | 2.45 | 9.86 | |
| **VCPA** | 3 | 33.41 | 2.84 | **51.19** | 2.46 | 9.48 | |
| **BIPLS** | 3 | 20.16 | 3.08 | **43.55** | 2.65 | 9.98 | |
|  | **Equilibrium Modulus (MPa)** | | | | |  |
| **MC-UVE-PLS** | 3 | 59.26 | 1.20 | **62.98** | 1.03 | 17.25 | |
| **GA-PLS** | 10 | 61.41 | 1.31 | **60.47** | 1.06 | 17.60 | |
| **VCPA** | 5 | 48.47 | 1.13 | **58.35** | 1.09 | 17.58 | |
| **JK-PLS** | 8 | 41.43 | 1.34 | **52.87** | 1.16 | 19.14 | |
| **CARS** | 2 | 44.49 | 1.26 | **45.37** | 1.25 | 19.82 | |
| **BIPLS** | 5 | 36.84 | 1.18 | **41.65** | 1.29 | 20.34 | |
|  | **Dynamic Modulus (MPa)** | | | | |  |
| **MC-UVE-PLS** | 3 | 63.53 | 4.80 | **72.31** | 3.23 | 12.93 | |
| **GA-PLS** | 3 | 50.88 | 4.83 | **69.75** | 3.38 | 12.29 | |
| **VCPA** | 2 | 43.06 | 4.86 | **67.61** | 3.50 | 13.76 | |
| **JK-PLS** | 5 | 43.64 | 5.27 | **67.47** | 3.50 | 13.49 | |
| **CARS** | 3 | 63.13 | 4.45 | **67.05** | 3.53 | 13.20 | |
| **BIPLS** | 7 | 38.73 | 4.85 | **55.98** | 4.07 | 14.87 | |

PLS models optimized by using variable selection methods show much better model performance (table 2) when compared to values in table 1. Significant improvement in the number of components is seen in table 2 as the models have lesser number of components than in PLSR of table 1. MC-UVE-PLS was found to improve the PLS performance parameters consistently.



**Figure 1**: Representative absorbance spectra of AC with different thickness values (0.84 mm, 0.71mm and 0.59mm) and 2nd derivative preprocessed spectra (top inset). The MC-UVE-PLS selection ranges (bottom inset) shows regions of wavelength selected (black bars) and empty spaces indicates the eliminated variables (white spaces).

The MC-UVE algorithm eliminates the wavelength variables (figure 1) depending on the relative importance of each variable in the calibration model.

# Discussion

In this study, for the first time a comparative analysis (table 1) of multivariate regression methods in NIR spectroscopy on cartilage data is presented. First, comparison of different optimized multivariate regression techniques, namely PLS, PCR, LASSO, ridge and LS-SVM, using key statistical parameters was conducted. PLS regression was found to be the best regression method based on this analysis.

Second, the effect of variable selection methods on the performance of PLS regression models was evaluated (table 2). The wavelength region of input NIR spectra was limited to 700-1050 nm[31] range to follow the procedure of the former study by Sarin et.al 2016. As a result, the variable selection methods had a relatively narrow spectrum of variables. Nonetheless, the results show the applicability of variable selection methods in regression analysis, which is consistent with the findings from Abrahamsson et.al[32] applying NIR transmission studies in intact tablets. The present result indicate that the variable selection improves the model performance and enhances the results of PLSR. In particular, MC-UVE-PLS is quite suitable in NIR spectroscopy of cartilage.

Recent cartilage research studies have favored projection regression methods such has PLSR and PCR due to ease of implementation. However, variable selection methods have not been applied earlier. PLSR modelled (table 1) the optical response of the tissue for articular cartilage thickness better (10% to 30% higher) compared to cartilage mechanical properties. This is attributed to direct relationship between the NIR spectra and the tissue thickness as the path length affects the light absorption as the rays traverses the tissue[7]. The regression comparison highlighted some limitations of sophisticated regression techniques in modelling cartilage data, as LS-LVM and PCR seemed to suffer from overfitting and under fitting, respectively. PLSR seemed to handle the multicollinearity in NIR data better and remain stable in comparison to other regression methods discussed, which is in agreement with the findings of Yeniay et.al[33].

Comparing the results of the current study and former study Sarin et al.[22] on the same equine data, promotes the usage of variable selection methods for cartilage thickness and dynamic modulus of cartilage tissue. MC-UVE-PLS prediction models displayed 7-8 % improvement in the *R2* test set and 15% lower RMSEP in case of thickness measurements and dynamic modulus respectively compared to former study. On the other hand, with instantaneous and equilibrium moduli, the variable selection methods showed no improvement in *R2* or RMSEP over the standard PLS models. This may be due to limited spectral range used in this study. However, it should be noted that the variable selection prediction models required lesser number of PLS components than the standard PLS models, which reduces the possibility of overfitting.

MC-UVE algorithm showed (figure 1) higher consistency in selection of variables in the 730 to 780nm and 925 to 980nm range. The NIR absorption spectra in AC arise mainly from CH, NH, OH and SH bonds[34] which form the elemental constituents of the cartilage matrix, and the information in this region indicates the micro- and macroscopic properties of the tissue condition. Cartilage NIR spectra in the 800 to 1100nm region is due to 3rd overtone of CH and NH bonds (PGs and collagen[35]), and 970nm is the OH bond associated with the water content of the tissue[36]. Hence, application of the MC-UVE method not only preserved the essential spectral information indicating the tissue condition, but also improved the performance of the prediction model.

The GA-PLS in NIR applications has been studied in horticultural studies[37], food engineering[38] and fuel analytical[39] studies as the optimal variable selection methods. However, in the current study on cartilage, MC-UVE-PLS surpassed GA-PLS and VCPA, also contrasting the results of Yun et.al study with VCPA[27]. Limitations imposed on the wavelength band might have reduced the performance of the VCPA, which in theory is the better variable selection method. Likewise, the narrow wavelength bandwidth probably limits the performance of all other variable selection methods, as there are less relevant variables to work with compared to the use of full NIR range. The results of the present study recommend using the PLS technique as the regression tool and that prediction of cartilage thickness and mechanical properties results can be further optimized by using MC-UVE-PLS.

# Acknowledgments

This study was funded by the Academy of Finland (project 267551, University of Eastern Finland) and Kuopio University Hospital (VTR projects 5041750 and 5041744, PY210 Clinical Neurophysiology).

# Contributions

# Prakash M: Analysis and interpretation of data and the main writer of the manuscript.

**Sarin, J.K.:** Acquisition of data and drafting of manuscript.

**Rieppo, L: C**onception of exploratory idea and drafting of manuscript**.**

**Afara I.O.:** Supervision of statistical analyses and drafting of manuscript.

**Töyräs J.:** Study conception and design, drafting of manuscriptand critical revision

All authors contributed in the preparation and approval of the final submitted manuscript.

# Conflict of Interest

The authors have no conflicts of interest in the execution of this study and preparation of the manuscript.

# References

[1] J.A. Buckwalter, H.J. Mankin, Articular cartilage: degeneration and osteoarthritis, repair, regeneration, and transplantation., Instr. Course Lect. 47 (1998) 487–504. http://europepmc.org/abstract/med/9571450 (accessed February 9, 2016).

[2] S. Saarakkala, P. Julkunen, P. Kiviranta, J. Mäkitalo, J.S. Jurvelin, R.K. Korhonen, Depth-wise progression of osteoarthritis in human articular cartilage: investigation of composition, structure and biomechanics, Osteoarthr. Cartil. 18 (2010) 73–81. doi:10.1016/j.joca.2009.08.003.

[3] F. Guilak, A. Ratcliffe, N. Lane, M.P. Rosenwasser, V.C. Mow, Mechanical and biochemical changes in the superficial zone of articular cartilage in canine experimental osteoarthritis, J. Orthop. Res. 12 (1994) 474–484. doi:10.1002/jor.1100120404.

[4] P.H. Puhakka, N.C.R. te Moller, I.O. Afara, J.T.A. Mäkelä, V. Tiitu, R.K. Korhonen, H. Brommer, T. Virén, J.S. Jurvelin, J. Töyräs, Estimation of articular cartilage properties using multivariate analysis of optical coherence tomography signal, Osteoarthr. Cartil. 23 (2015) 2206–2213. doi:10.1016/j.joca.2015.05.034.

[5] L. Rieppo, J. Rieppo, J.S. Jurvelin, S. Saarakkala, Fourier transform infrared spectroscopic imaging and multivariate regression for prediction of proteoglycan content of articular cartilage, PLoS One. 7 (2012). doi:10.1371/journal.pone.0032344.

[6] A. Bonifacio, C. Beleites, F. Vittur, E. Marsich, S. Semeraro, S. Paoletti, V. Sergo, Chemical imaging of articular cartilage sections with Raman mapping, employing uni- and multi-variate methods for data analysis., Analyst. 135 (2010) 3193–3204. doi:10.1039/c0an00459f.

[7] I. Afara, S. Singh, A. Oloyede, Application of near infrared (NIR) spectroscopy for determining the thickness of articular cartilage, Med. Eng. Phys. 35 (2013) 88–95. doi:10.1016/j.medengphy.2012.04.003.

[8] U.P. Palukuru, C.M. McGoverin, N. Pleshko, Assessment of hyaline cartilage matrix composition using near infrared spectroscopy, Matrix Biol. 38 (2014) 3–11. doi:10.1016/j.matbio.2014.07.007.

[9] C.M. McGoverin, A. Hanifi, U.P. Palukuru, F. Yousefi, P.B.M. Glenn, M. Shockley, R.G. Spencer, N. Pleshko, Non-destructive Assessment of Engineered Cartilage Composition by Near Infrared Spectroscopy., Ann. Biomed. Eng. (2016). doi:10.1007/s10439-015-1536-8.

[10] G. Spahn, H. Plettenberg, H. Nagel, E. Kahl, H.M. Klinger, T. Mückley, M. Günther, G.O. Hofmann, J.A. Mollenhauer, Evaluation of cartilage defects with near-infrared spectroscopy (NIR): an ex vivo study., Med. Eng. Phys. 30 (2008) 285–92. doi:10.1016/j.medengphy.2007.04.009.

[11] M. V. Padalkar, R.G. Spencer, N. Pleshko, Near infrared spectroscopic evaluation of water in hyaline cartilage, Ann. Biomed. Eng. 41 (2013) 2426–2436. doi:10.1007/s10439-013-0844-0.

[12] L.P.H. Kim H. Esbensen, Dominique Guyot, Frank Westad, Multivariate Data Analysis: In Practice : an Introduction to Multivariate Data Analysis and Experimental Design, 2002.

[13] T. Næs, H. Martens, Principal component regression in NIR analysis: Viewpoints, background details and selection of components, J. Chemom. 2 (1988) 155–167. doi:10.1002/cem.1180020207.

[14] J. Trygg, S. Wold, PLS regression on wavelet compressed NIR spectra, Chemom. Intell. Lab. Syst. 42 (1998) 209–220. doi:10.1016/S0169-7439(98)00013-6.

[15] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, J. R. Stat. Soc. 58 (1996) 267–88.

[16] F. Chauchard, R. Cogdill, S. Roussel, J.M. Roger, V. Bellon-Maurel, Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes, Chemom. Intell. Lab. Syst. 71 (2004) 141–150. doi:10.1016/j.chemolab.2004.01.003.

[17] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-infrared spectroscopy, Anal. Chim. Acta. 667 (2010) 14–32. doi:10.1016/j.aca.2010.03.048.

[18] F. Westad, H. Martens, Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression, J. Near Infrared Spectrosc. 8 (2000) 117. doi:10.1255/jnirs.271.

[19] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in Partial Least Squares Regression, Chemom. Intell. Lab. Syst. 118 (2012) 62–69. doi:10.1016/j.chemolab.2012.07.010.

[20] R.. Korhonen, M.. Laasanen, J. Töyräs, J. Rieppo, J. Hirvonen, H.. Helminen, J.. Jurvelin, Comparison of the equilibrium response of articular cartilage in unconfined compression, confined compression and indentation, J. Biomech. 35 (2002) 903–909. doi:10.1016/S0021-9290(02)00052-0.

[21] A.M. Fard, P. Vacas-Jacques, E. Hamidi, H. Wang, R.W. Carruth, J.A. Gardecki, G.J. Tearney, Optical coherence tomography--near infrared spectroscopy system and catheter for intravascular imaging., Opt. Express. 21 (2013) 30849–58. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3926541&tool=pmcentrez&rendertype=abstract (accessed March 17, 2016).

[22] J.K. Sarin, M. Amissah, H. Brommer, D. Argüelles, J. Töyräs, I.O. Afara, Near Infrared Spectroscopic Mapping of Functional Properties of Equine Articular Cartilage, Ann. Biomed. Eng. (2016) 1–11. doi:10.1007/s10439-016-1659-6.

[23] J. Reunanen, Overfitting in Making Comparisons Between Variable Selection Methods, J. Mach. Learn. Res. 3 (2003) 1371–1382.

[24] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, WORLD SCIENTIFIC, 2002. doi:10.1142/5089.

[25] H. Li, Y. Liang, Q. Xu, D. Cao, Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration, Anal. Chim. Acta. 648 (2009) 77–84. doi:10.1016/j.aca.2009.06.046.

[26] L.Y.-Z. Li H.-D., Xu Q.-S., An Integrated Library for Partial Least Squares Regression and Discriminant Analysis, PeerJ Prepr. 2e190v. (2014).

[27] Y.-H. Yun, W.-T. Wang, B.-C. Deng, G.-B. Lai, X. Liu, D.-B. Ren, Y.-Z. Liang, W. Fan, Q.-S. Xu, Using variable combination population analysis for variable selection in multivariate calibration, Anal. Chim. Acta. 862 (2015) 14–23. doi:10.1016/j.aca.2014.12.048.

[28] R. Leardi and A. Lupiáñez, Genetic algorithms applied to feature selection in PLS regression: how and when to use them, Chemom. Intell. Lab. Syst. 41 (1998) 95–207.

[29] B. Efron, The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, 1982. doi:10.1137/1.9781611970319.

[30] I. Afara, I. Prasadam, R. Crawford, Y. Xiao, a. Oloyede, Non-destructive evaluation of articular cartilage defects using near-infrared (NIR) spectroscopy in osteoarthritic rat models and its direct relation to Mankin score, Osteoarthr. Cartil. 20 (2012) 1367–1373. doi:10.1016/j.joca.2012.07.007.

[31] M. V Padalkar, N. Pleshko, Wavelength-dependent penetration depth of near infrared radiation into cartilage., Analyst. 140 (2015) 2093–100. doi:10.1039/c4an01987c.

[32] C. Abrahamsson, J. Johansson, A. Sparén, F. Lindgren, Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets, (n.d.). doi:10.1016/S0169-7439(03)00064-9.

[33] O. Yeniay, A. Goktas, A comparison of partial least squares regression with other prediction methods, Hacettepe J. Math. (2002). http://www.hjms.hacettepe.edu.tr/uploads/ce7bdb8f-5f89-4f03-9822-a20e2ee35d03.pdf (accessed August 25, 2016).

[34] I. Afara, T.R. Sahama, A. Oloyede, Near infrared for non-destructive testing of articular cartilage, Int. Symp. Nondestruct. Test. Mater. Struct. 6 (2011) 399–404. doi:10.1007/978-94-007-0723-8\_58.

[35] M. Ghiassi-Nejad, P.. Torzilli, H. Peemoeller, M.. Pintar, Proton spin–spin relaxation study of molecular dynamics and proteoglycan hydration in articular cartilage, Biomaterials. 21 (2000) 2089–2095. doi:10.1016/S0142-9612(00)00140-X.

[36] P.A. Torzilli, R. Grigiene, J. Borrelli, D.L. Helfet, Effect of Impact Load on Articular Cartilage: Cell Metabolism and Viability, and Matrix Water Content, J. Biomech. Eng. 121 (1999) 433. doi:10.1115/1.2835070.

[37] Z. Xiaobo, Z. Jiewen, H. Xingyi, L. Yanxiao, Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of “Fuji” apple based on different PLS models, Chemom. Intell. Lab. Syst. 87 (2007) 43–51. doi:10.1016/j.chemolab.2006.09.003.

[38] Y. Ying, Y. Liu, Nondestructive measurement of internal quality in pear using genetic algorithms and FT-NIR spectroscopy, J. Food Eng. 84 (2008) 206–213. doi:10.1016/j.jfoodeng.2007.05.012.

[39] M.C. Breitkreitz, I.M. Raimundo, Jr, J.J.R. Rohwedder, C. Pasquini, H.A. Dantas Filho, G.E. José, M.C.U. Araújo, Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration, Analyst. 128 (2003) 1204–1207. doi:10.1039/B305265F.