Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 00 (2018) 000–000

Energy Procedia

www.elsevier.com/locate/procedia

10th International Conference on Applied Energy (ICAE2018), 22-25 August 2018, Hong Kong, China

Real-time Biomass Characterization in Energy Conversion Processes using Near Infrared Spectroscopy - A Machine Learning Approach

Mobyen Uddin Ahmed[a], Peter Andersson[a], Tim Andersson[a], Elena Tomas Aparicio[b,c], Hampus Baaz[a], Shaibal Barua[a,d], Albert Bergström[a], Daniel Bengtsson[a], Jan Skvaril[b,*], and Jesús Zambrano[b]

[a]*School of Innovation, Design and Engineering, Mälardalen University, SE-72123 Västerås, Sweden*
[b]*Future Energy Center, School of Business, Society and Engineering, Mälardalen University, SE-72123 Västerås, Sweden*
[c]*Mälarenergi AB, Sjöhagsvägen 3, Västerås 721 03, Sweden*
[d]*RISE SICS Västerås, Stora Gatan 36, Västerås 722 12, Sweden*

Abstract

The aim of this work is to apply and evaluate different chemometric approaches employing several machine learning techniques in order to characterize the moisture content in biomass from data obtained by Near Infrared (NIR) spectroscopy. The approaches include three main parts*: a) data pre-processing, b) wavelength selection* and *c) development of a regression model* enabling moisture content measurement. Standard Normal Variate (SNV), Multiplicative Scatter Correction and Savitzky-Golay first (SG_1) and second (SG_2) derivatives and its combinations were applied for data pre-processing. Genetic algorithm (GA) and iterative PLS (iPLS) were used for wavelength selection. Artificial Neural Network (ANN), Gaussian Process Regression (GPR), Support Vector Regression (SVR) and traditional Partial Least Squares (PLS) regression, were employed as machine learning regression methods. Results shows that SNV combined with SG1 first derivative performs the best in data pre-processing. The GA is the most effective methods for variable selection and GPR achieved a high accuracy in regression modeling while having low demands on computation time. Overall, the machine learning techniques demonstrate a great potential to be used in future NIR spectroscopy applications.

Selection and peer-review under responsibility of the scientific committee of the 10th International Conference on Applied Energy (ICAE2018).

Keywords: Artificial Neural Netwrok; Chemometrics; Gaussian Process Regression; Multiplicative Scatter Correction; Standard Normal Variate; Support Vector Regression; Partial Least Squares; Savitzky-Golay derivatives.

[*] Corresponding author. Tel.: +46 736 620977.
E-mail address: jan.skvaril@mdh.se

Selection and peer-review under responsibility of the scientific committee of the 10th International Conference on Applied Energy (ICAE2018).

1. Introduction

Nowadays, society strives to increase the share of renewable and alternative energy sources with motivation to reduce dependence on fossil fuels and to reduce the emissions of carbon dioxide. Biomass is considered as the only carbon-based sustainable solution to replace fossil-based fuels. However, biomass is characterized by strong physical and chemical diversity, which makes it energy utilization challenging. Since the energy biomass conversion processes are sensitive to high variability in feedstock material properties (such as moisture content) and requires continuous regulation, it is needed a non-destructive method able to measure biomass in real-time [1].

As demonstrated in previous studies, real-time measurements can be achieved by employing near infrared (NIR) spectroscopy. This technique is based on interactions between emitted electromagnetic radiation (EMR) and the analyzed material. Such interactions lead to vibrational transitions in structural molecular groups e.g. O-H, C-H, N-H, S-H, C=O, C=H and C=C, producing measurable/detectable response in the spectra which is, according to Beer–Lambert law, linearly proportional to the concentration of the absorbing molecule. The extraction of the information from the spectra is done by a chemometric approach including mathematical and statistical methods to provide maximum chemical information [1]. The chemometric approach usually consists of spectral pre-processing, wavelength selection an employment of regression or discrimination method. The state-of-practice in the field is the use of multivariate liner techniques, e.g. Partial Least Squares (PLS) regression. However, the recent literature review has identified emerging opportunity for application of other mostly non-linear machine learning methods [2].

In the literature different pre-processing, wavelength selection and multivariate calibration methods are proposed. The combination or exclusion of the methods depends on the material as well as its physical and chemical properties. For example, marginal difference have been observed comparing raw and pre-processed NIR dataset used for determination of moisture content in marzipan [3], whereas [4-6] have argued that major improvements could be done by applying pre-processing methods. Comparison of different wavelength selection methods can be found in [7]. Wavelength selection may be applied, because only a fraction of the NIRS data from the waves of the NIR spectrum is affected upon moisture variation in the material [3]. By performing wavelength selection, a lot of data may be neglected, therefore computations will become faster and prediction results may improve [4]. Regression models are created with multivariate calibration methods that are trained using supervised learning. Different methods vary in speed, accuracy, complexity and application [8]. PLS is the most common used method [4-6, 9-11], however non-linear method such as GPR is present in some of the recent studies.

The objective of the present work is to evaluate various chemometric approaches using machine-learning methods for biomass moisture determination by NIR spectroscopy. The study includes acquisition of NIR spectral data and reference biomass moisture determination according to standardized laboratory method. Most importantly, the complex chemometric approaches consisting of individual methods for data pre-processing, wavelength selection and machine-learning regression are employed and compared according to proposed evaluation methodology. The results are presented in the form of Coefficient of determination (R^2), Root Mean Squared Error computed from the selected cross validation round (RMSECV) and standard deviation (s) [12].

2. Materials and Methods

2.1. Dataset

Dataset was obtained by experimental measurements on solid biomass samples (random blends of pine and spruce wood chips, bark, forest residues and sawdust, particle size approx. 5-50 mm) collected from biomass processing facilities in Västmanland region, Sweden. Spectral data were acquired using a FT-NIR spectrometer MATRIX-F equipped with contactless illumination/detection head Q410A (both Bruker Optics, Germany) with 4 halogen sources [13]. Near infrared spectra was collected in diffuse reflection mode on samples moving on a turntable at 1 $m \cdot s^{-1}$ and recorded as relative absorbance [2]. Acquisition parameters are summarized in Table 1.

Table 1. Acquisition parameters of NIR spectra

Parameter	**Value**
Focal distance of illumination/detection head	170 mm
Scanning spot size	approx. ø10 mm
Recorded spectral range	12000 – 4000 cm^{-1} (834-2500 nm)
Number of averaged scans	32
Spectral resolution	16 cm^{-1}
Scanner velocity	10 kHz
Ambient temperature	20±1 °C
Number of data points in each spectra	1037
Number of tested samples	809

The determination of reference value (i.e. fuel moisture content) was carried out according to standardized laboratory procedure EN ISO 18134-2:2015 [14]. The method is based on thermo-gravimetric measurements where samples are thermally treated at 105° C.

2.2. Chemometric approach

The process of moisture contents prediction in biomass consisted of three phases: pre-processing of the NIR data, wavelength selection and development of a regression model. The process is shown in Fig 1.

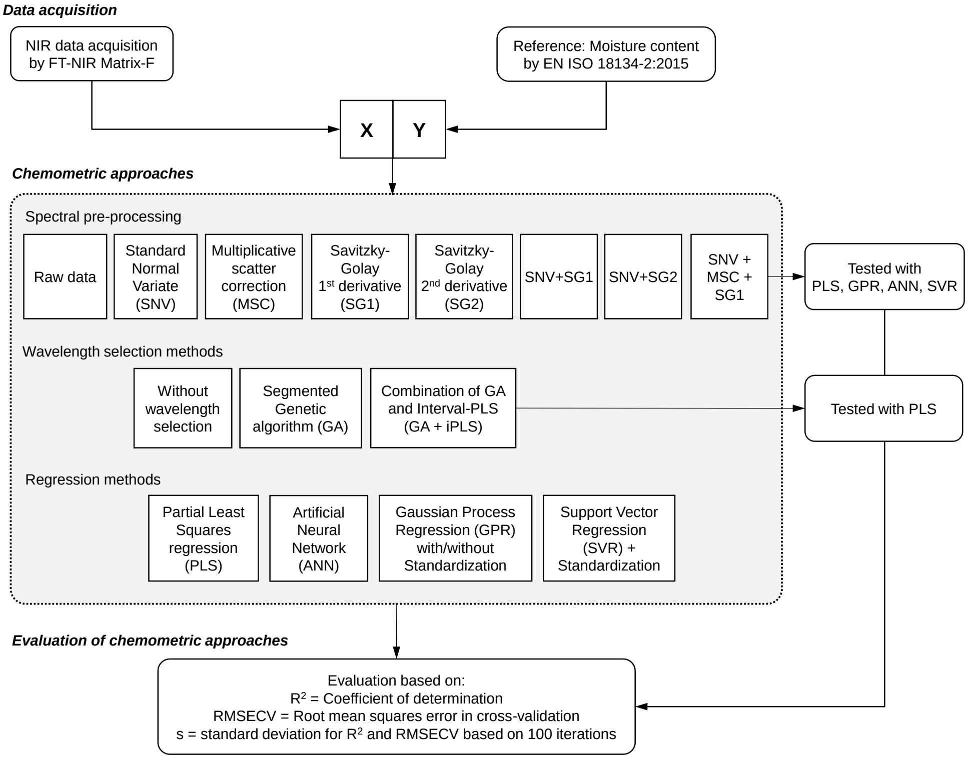
Data acquisition
NIR data acquisition by FT-NIR Matrix-F
Reference: Moisture content by EN ISO 18134-2:2015
X
Y
Chemometric approaches
Spectral pre-processing
Raw data
Standard Normal Variate (SNV)
Multiplicative scatter correction (MSC)
Savitzky-Golay 1st derivative (SG1)
Savitzky-Golay 2nd derivative (SG2)
SNV+SG1
SNV+SG2
SNV + MSC + SG1
Tested with PLS, GPR, ANN, SVR
Wavelength selection methods
Without wavelength selection
Segmented Genetic algorithm (GA)
Combination of GA and Interval-PLS (GA + iPLS)
Tested with PLS
Regression methods
Partial Least Squares regression (PLS)
Artificial Neural Network (ANN)
Gaussian Process Regression (GPR) with/without Standardization
Support Vector Regression (SVR) + Standardization
Evaluation of chemometric approaches
Evaluation based on:
R2 = Coefficient of determination
RMSECV = Root mean squares error in cross-validation
s = standard deviation for R2 and RMSECV based on 100 iterations

Fig. 1. Chemometric approach used in our study.

Three separate pre-processing methods are used to generate three sets of pre-processed datasets, these are: Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and Savitzky-Golay derivatives (SG) with both first and second derivative (SG_1 and SG_2, respectively). Each of these pre-processing methods was evaluated employing Partial Least Square regression (PLS), Gaussian process regression (GPR), Artificial Neural Network (ANN) and Support Vector Regression (SVR). The PLS was tested while including 17-20 principle components; the GPR was tested with an exponential-kernel function; the ANN was tested with Bayesian Regulated back propagation with one hidden layer and 9 nodes; and the SVR was tested with a Gaussian-kernel function. In this work, all programs were written in the proprietary programming language MATLAB[1].

Genetic algorithm (GA) [7] in combination with reverse interval-PLS (iPLS) [15] is used for wavelength selection. The GA was used to select minimum wavelengths, i.e. wavelengths that contain relevant information are frequently selected alongside maximizing the prediction accuracy. The iPLS divides the spectrum of 1037 wavelengths into 21 arbitrary intervals of equal size with 49 wavelengths per interval. The algorithm removes one interval at each iteration and evaluates in term of improvement of RMSECV and R^2. This was performed 150 times for full-iPLS and 2000 for 21 divided-iPLS to create a sorted vector of the most used intervals. The best intervals are decided by “Wavelength selection” and then tested using GA to specify specific wavelengths. The last few wavelengths which did not fit into an interval are sent through and evaluated by the GA when it evaluates the intervals afterwards. When the intervals are known, the GA will select the individual wavelengths from these intervals that give the best accuracy. All the wavelength selection methods have been tested 100 times to find the most frequently selected wavelengths. Each generation of GA was set to consist of 30 individuals, where an individual is represented by a bit-vector. The population was randomly initialized for the first generation and the other generations had a high grade of elitism such that 70% of the next generation consisted of the best individuals from the previous generation. The remaining 30% were the offspring’s from parents selected accordingly to a Roulette Wheel Selection scheme [16] and the crossover was done with the Uniform method. When the maximum number of generations has been reached, a new test will begin. This will be repeated 100 times where the algorithm will keep track of how often the individual wavelengths were used in all the tests together.

Finally, when the optimal wavelength intervals are identified, the best regression methods based on the prediction performance with pre-processing methods are separately used so that result could be further improved with respective selected wavelengths or at least to keep the same accuracy but with less wavelengths.

2.3. Evaluation method

Root Mean Square Error (RMSE) and coefficient of determination (R^2) were used to evaluate the performance of the regression method in term of prediction accuracy, defined as

$$RMSE = \sqrt{\frac{\sum(\hat{e}-e)^2}{m}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_1^m (e_i-\hat{e}_i)^2}{\sum_1^m (e_i-\bar{e}_i)^2} \quad (2)$$

RMSE is a popular method for measuring prediction by taking the root of the squared average difference between the predicted value $\hat{e}$ generated by a model and the actual value e across m samples, the amount of difference between $\hat{e}$ and e is proportional to the result. This follows that larger errors affect the RMSE exponentially compared to small errors. Models that regularly predict close to the targeted value score a lower RMSE compared to models that make perfect prediction. The variation of the response variable can be measured by R^2. In addition, to avoid model overfitting, the holdout method was applied [17], which partitioned samples into training and test dataset, where

[1] MATLAB Version: 9.3.0.713579 (R2017b).

training dataset consisted of 70% and testing dataset consisted of 30% of entire dataset. The training set was used to fit a model, where the performance was evaluated based on the prediction of the fitted model on the test dataset. Moreover, the training was validated with 6-fold cross-validation. Hence, RMSECV was estimated for each regression model with training dataset and another evaluation criterion was computation time of training a model along with data pre-processing and wavelength selection.

3. Results and Discussion

Complex evaluation have been performed in order to examine the combination of pre-processing, wavelength selection and regression methods. Table 2 shows the performance of PLS and GPR using pre-processing methods on training dataset, where all wavelengths variables of NIR were used for predicting moisture content. The SNV and MSC provided very similar results over raw data with an improvement of 6-33% for SNV and 1-33% for MSC depending of the regression method. MSC required considerable more time to prepare the provided data compared to SNV. Similar results were also observed when SNV and MSC were used in combination with the SG_1. SNV combined with SG_1 was the pre-processing method with the best performance improvement considering also the computation time. GPR even performed better when MATLAB's built-in standardization method[†] was used.

Table 2. Result of the pre-process evaluation in terms of RMSECV and execution time on training dataset. Note that the computing times are on the whole data set, i.e. they are just a comparison and do not show the actual time taken in real-time use. Standard deviation (s) is estimated over 6-fold cross-validation of 100 iterations.

Dataset	Execution time (ms , s)	PLS		GPR		ANN		SVR	
		(RMSECV, s)	(R^2, s)	(RMSECV, s)	(R^2, s)	(RMSECV, s)	(R^2, s)	(RMSECV, s)	(R^2, s)
Raw data	NaN, NaN	2.73, 0.12	0.97, 0.003	3.0, $s = 0.18$	0.95, 0.003	2.31, 0.10	0.93, 0.003	2.47, 0.15	0.94, 0.003
SNV	24.5, 0.14	2.41, 0.09	0.98, 0.002	2.0, $s = 0.14$	0.98, 0.002	2.16, 0.10	0.98, 0.002	2.14, 0.12	0.97, 0.002
MSC	796.1, 43.7	2.46, 0.09	0.98, 0.002	2.01, 0.14	0.97, 0.002	2.32, 0.18	0.98, 0.002	2.15, 0.07	0.97, 0.002
SG_1	60.1, 6.0	2.75, 0.11	0.97, 0.003	2.23, 0.13	0.97, 0.002	2.42, 0.15	0.98, 0.002	2.17, 0.12	0.97, 0.002
SG_2	59.2, 2.6	2.88, 0.12	0.97, 0.003	2.19, 0.09	0.96, 0.002	2.67, 0.11	0.97, 0.002	2.24, 0.27	0.96, 0.003
SNV+SG_1	85.1, 7.4	2.31, 0.09	0.98, 0.002	2.03, 0.09	0.98, 0.002	2.26, 0.08	0.98, 0.002	2.04, 0.09	0.98, 0.002
MSC+SG_2	841.5, 17.2	2.36, 0.08	0.98, 0.002	2.04, 0.09	0.98, 0.002	2.34, 0.10	0.98, 0.002	2.03, 0.08	0.98, 0.002
SNV+MSC+SG_1	869.2, 42.8	2.35, 0.08	0.98, 0.002	2.03, 0.09	0.98, 0.002	2.28, 0.11	0.98, 0.002	2.04, 0.06	0.98, 0.002

Each method was evaluated with test dataset and their best respective pre-processing method, according to Table 2. The best pre-processing method was evaluated considering both accuracy and speed. Table 3 represents the evaluation of the regression models on test dataset. It clearly shows that the GPR has the lowest RMSECV. Note that the computation time covers the whole test set and it includes pre-processing, followed by regression through the set.

Table 3. Evaluation of regression methods with optimal setups on the full spectrum. Comparison between the two methods in term of execution time, RSME and R^2. Average execution time, RSME and R^2 are presented; both the mean value and standard deviation (s) are estimated over 100 iterations.

Dataset	Execution time (ms, s)	RMSECV, s	R^2, s
SNV+SG_1+PLS	20.89, 0.73	2.31, 0.09	0.98, 0.002
SNV+SG_1+ GPR	33.62, 1.47	1.69, 0.10	0.98, 0.001

[†] https://se.mathworks.com/help/stats/fitrgp.html

SNV + ANN	19.60, 1.25	2.01, 0.12	0.98, 0.002
SNV + SG_1 + SVR	36.27, 2.53	1.96, 0.14	0.98, 0.002

The similarities between SNV and MSC are very understandable as both methods are used for the same purpose, light scatter correction and adjusting spectral baseline [18]. SG_1 and SG_2 are methods for eliminating outliers in spectral baseline variation between samples while enhancing small differences [19]. This could explain why SG_1 and SG_2 performed poorly when used on their own. Background noise in the acquired spectra could possibly have been enhanced by SG_1 and SG_2. This also explains the improvement when SG_1 and SG_2 were combined with SNV and MSC, as these methods reduce background noise.

In the optimal wavelength selection process, data pre-processing was included with regression method if the wavelength selection is run with pre-processed data, otherwise wavelet selection was done on raw dataset. The biggest difference is observed in the wavelengths over 2000 nm in almost all the different methods. Other than that, most wavelength intervals are the same, with some small offsets. Tables 4 shows the RMSECV, QTY (the number of wavelengths used) and the distribution of wavelengths in the spectrum, with a percentage (%) of QTY, for each wavelength selection method. All the suggested wavelengths have been evaluated with PLS and GPR. The GA performed better on the full NIR-spectrum with the best suited pre-processing since it only uses input of 71 wavelength data points and still achieves a prediction accuracy of 1.64% RMSECV. Reduction of input wavelength data points positively contribute to model robustness. Fig 2 shows the corresponding results for the GPR model with best accuracy on test dataset using raw dataset and pre-processed dataset.

Table 4. Moisture contents prediction using PLS and GPR with wavelet selection. QTY refers to number of wavelengths used.

Goal	Raw dataset					Pre-processed dataset				
	Wavelength used		Average RMSECV	R^2	QTY	Wavelength used		Average RMSECV	R^2	QTY
	nm	%				nm	%			
Least number of wavelengths to keep the same accuracy as using all wavelengths in the segment	844–914 968–993 1143–1202 1353–1411 1476 1649–1653 1884 1952	19 25 22 16 3 9 3 3	PLS: 2.75, $s = 0.11$ GPR: 2.95, $s = 0.23$	PLS: 0.96, $s = 0.002$ GPR: 0.96, $s = 0.002$	32	1073–1075 1143–1202 1346–1447 1538 1620–1631 1753–1803 1865–2016 2087 2175–2361	25 4 21 1 8 8 16 1 16	PLS: 3.06% $s = 0.08$ GPR: 1.64% $s = 0.12$	PLS: 0.98 $s = 0.002$ GPR: 0.98 $s = 0.002$	71
Best accuracy	834–1204 1272–1301 1352–1497 1590–1594 1645–1653 1719–1773 1851–2084 2136–2179 2254–2270 2353	44 3 20 2 3 3 20 2 2 1	PLS: 2.45, $s = 0.07$ GPR: 2.81, $s = 0.21$	PLS: 0.96, $s = 0.003$ GPR: 0.96, $s = 0.002$	147	834–1656 1706–2427	62 38	PLS: 2.13% $s = 0.07$ GPR: 1.68% $s = 0.09$	PLS: 0.98 $s = 0.002$ GPR: 0.98 $s = 0.002$	355

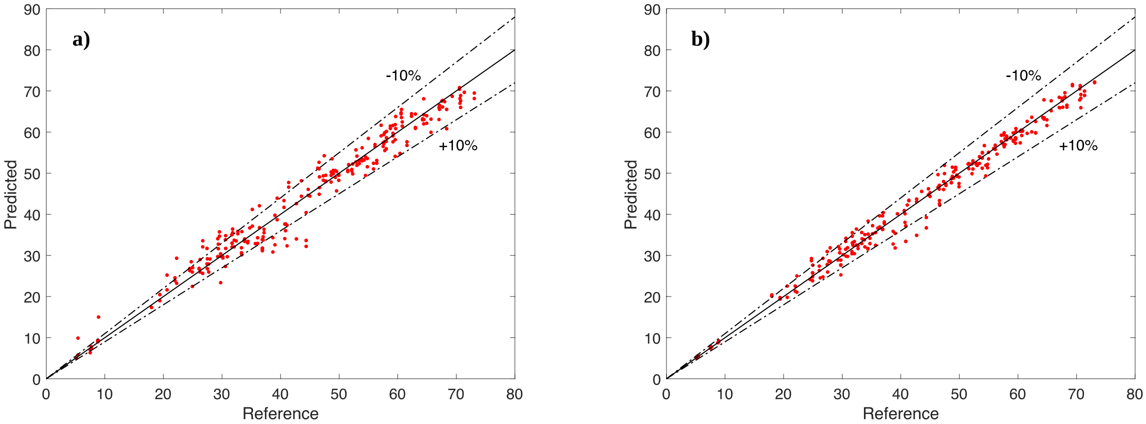

Fig. 2. Resulting Parity plots on test data using GPR. a) Parity plot shows the best prediction with wavelengths from raw dataset; b) Parity plot shows the best prediction with selected wavelengths from pre-processed dataset

Depending on the method applied, different wavelengths were selected. Some regions were more frequently selected by all the methods. It has been observed that the most frequently selected wavelengths correlates, in most of the cases, with the optical absorption features of water molecules. Absorption features arises from vibrational transitions – fundamental vibrations and overtones and combinations: ν_1 (H-O-H symmetric stretching transition), ν_2 (H-O-H bending mode transition) and ν_3 (H-O-H asymmetric stretching transition). The absorption feature at approx. 950 nm is assigned to $2\nu_1+\nu_3$, at approx.1280 nm to a $\nu_1+\nu_2+\nu_3$, at approx. 1400 nm to a $\nu_1+\nu_3$ and the one at 1900 nm to $\nu_2+\nu_3$ combination of vibrational transitions [20]. However, some of the selected wavelengths do not fall into the known absorption bands for water molecules. One reason for this could be that water molecules interacts with other substances in the material [7]. Some of the tests also used pre-processing, which have manipulated the data. This might also affect the wavelengths selected in a test set. The evaluation of the pre-processing methods shows that the faster pre-processing methods also provided a best accuracy, which is highly advantageous, as a low computation time is desired for the application in real-time. The results shows that the GPR achieves significantly greater accuracy compare to traditional PLS while having similar demands on computation time.

4. Conclusion

This paper presents evaluation of different chemometric approaches combining pre-processing, wavelength selection and machine-learning regression methods for biomass moisture characterization by near infrared (NIR) spectroscopy. According to the results presented, the following conclusions can be drawn:

- Application of pre-processing techniques to the NIR spectral data improved the results greatly in all of the cases compared to raw data. Hence, Multiplicative scatter correction (MSC) required considerably more time than other techniques. For given application, Standard Normal Variate (SNV) combined with Savitzky-Golay first derivative (SG_1) is the method with the best potential to improve model performance considering also the computation time.
- From evaluated methods, the Genetic Algorithm (GA) is the best performing wavelength selection method. It leads to significant reduction of input wavelength data points positively affecting model robustness while having no negative impact on model accuracy.
- Gaussian process regression (GPR) is considered the best performing machine-learning method for given dataset. The method achieved high accuracy while having relatively low demands on computation time

compared to other methods.

- Overall, the machine learning methods used in this work demonstrate great potential to substitute traditional methods as PLS regression and to become an important part of chemometric approaches in future NIR spectroscopy applications.

Acknowledgements

The research leading to these results has received funding from EU Horizon 2020 SPIRE-2 under the project FUDIPO. Jan Skvaril has also received funding from Swedish Energy Agency - Energimyndigheten under the project OPtiC-NIRS. In addition, the authors would like to acknowledge assistance from Daniele Orisio for performing part of NIR scans and reference laboratory measurements.

References

[1] J. Skvaril, K. Kyprianidis, A. Avelin, M. Odlare, E. Dahlquist, Fast Determination of Fuel Properties in Solid Biofuel Mixtures by Near Infrared Spectroscopy, Energy Procedia, 105 (2017) 1309-1317.
[2] J. Skvaril, K.G. Kyprianidis, E. Dahlquist, Applications of near-infrared spectroscopy (NIRS) in biomass energy conversion processes: A review, Applied Spectroscopy Reviews, 52 (2017) 675-728.
[3] Å. Rinnan, F.v.d. Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends in Analytical Chemistry, 28 (2009) 1201-1222.
[4] G. Kim, S.-J. Hong, A.-Y. Lee, Y.-E. Lee, S. Im, Moisture Content Measurement of Broadleaf Litters Using Near-Infrared Spectroscopy Technique, Remote Sensing, 9 (2017) 1212.
[5] B. Leblon, O. Adedipe, G. Hans, A. Haddadi, S. Tsuchikawa, J. Burger, R. Stirling, Z. Pirouz, K. Groves, J. Nader, A. LaRocque, A review of near-infrared spectroscopy for monitoring moisture content and density of solid wood, The Forestry Chronicle, 89 (2013) 595-606.
[6] T. Lestander, Multivariate NIR Studies of Seed-Water Interaction in Scots Pine Seeds (Pinus sylvestris L.), 2011.
[7] T.A. Lestander, R. Leardi, P. Geladi, Selection of near Infrared Wavelengths Using Genetic Algorithms for the Determination of Seed Moisture Content, Journal of Near Infrared Spectroscopy, 11 (2003) 433-446.
[8] L. Barry, A user-friendly guide to multivariate calibration and classification, Tomas Naes, Tomas Isakson, Tom Fearn and Tony Davies, NIR Publications, Chichester, 2002, ISBN 0-9528666-2-5, £45.00, Journal of Chemometrics, 17 (2003) 571-572.
[9] W. Ni, L. Nørgaard, M. Mørup, Non-linear calibration models for near infrared spectroscopy, Analytica Chimica Acta, 813 (2014) 1-14.
[10] A. Borin, M.F. Ferrão, C. Mello, D.A. Maretto, R.J. Poppi, Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk, Analytica Chimica Acta, 579 (2006) 25-32.
[11] L.-j. Xie, Y.-b. Ying, Use of near-infrared spectroscopy and least-squares support vector machine to determine quality change of tomato juice, Journal of Zhejiang University SCIENCE B, 10 (2009) 465-471.
[12] J.M. Bland, D.G. Altman, Statistics Notes: Measurement error, BMJ, 313 (1996) 744.
[13] J. Skvaril, Near-Infrared Spectroscopy and Extractive Probe Sampling for Biomass and Combustion Characterization, in, Mälardalen University, 2017.
[14] EN ISO 18134–2. 2015. Solid biofuels – Determination of moisture content – Oven dry method – Part 2: Total moisture – Simplified method., in.
[15] L. Riccardo, N. Lars, Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions, Journal of Chemometrics, 18 (2004) 486-497.
[16] N. Mohd Razali, J. Geraghty, Genetic Algorithm Performance with Different Selection Strategies in Solving TSP, 2011.
[17] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in:

Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2, Morgan Kaufmann Publishers Inc., Montreal, Quebec, Canada, 1995, pp. 1137-1143.

[18] S. Romero-Torres, J.D. Pérez-Ramos, K.R. Morris, E.R. Grant, Raman spectroscopic measurement of tablet-to-tablet coating variability, Journal of Pharmaceutical and Biomedical Analysis, 38 (2005) 270-274.

[19] Y. Lai, Y. Ni, S. Kokot, Discrimination of Rhizoma Corydalis from two sources by near-infrared spectroscopy supported by the wavelet transform and least-squares support vector machine methods, Vibrational Spectroscopy, 56 (2011) 154-160.

[20] S. Jacquemoud, S. Ustin, Application of radiative transfer models to moisture content estimation and burned land mapping, in: 4th International Workshop on Remote Sensing and GIS Applications to Forest Fire Management, 2003.