CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Rapid quantification of NaDCC for water purification tablets in commercial production using ATR-FTIR spectroscopy based on machine learning techniques
Authors	Asadi, Hamzeh;O'Mahony, Tom;Lambert, Julie;Brown, Kenneth N.
Publication date	2023-02-23
Original Citation	Asadi, H., O'Mahony, T., Lambert, J. and Brown, K.N. (2023) 'Rapid quantification of nadcc for water purification tablets in commercial production using atr-ftir spectroscopy based on machine learning techniques', AICS 2022, in L. Longo and R. O'Reilly (eds) Artificial Intelligence and Cognitive Science, Communications in Computer and Information Science, vol 1662, Cham: Springer Nature Switzerland, pp. 106–120. https://doi.org/10.1007/978-3-031-26438-2_9
Type of publication	book-chapter;Conference item
Link to publisher's version	10.1007/978-3-031-26438-2_9
Rights	© 2023 The Author(s). Open Access. This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made - http://creativecommons.org/licenses/by/4.0/
Download date	2026-11-17 06:33:00
Item downloaded from	https://hdl.handle.net/10468/14348



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh



Rapid Quantification of NaDCC for Water Purification Tablets in Commercial Production Using ATR-FTIR Spectroscopy Based on Machine Learning Techniques

Hamzeh Asadi[1(✉)], Tom O'Mahony[2], Julie Lambert[2], and Kenneth N. Brown[1]

[1] School of Computer Science and Information Technology, University College Cork, Cork, Ireland
{hamze.asadi,k.brown}@cs.ucc.ie
[2] Medentech Limited, Wexford, Ireland
{tom.omahony,julie.lambert}@kersia-group.com

Abstract. Accurate, fast and simple quantitative analysis of solid dosage forms is required for efficient pharmaceutical manufacturing. A spectroscopic analysis in ATR-FTIR (Attenuated Total Reflection-Fourier Transform Infrared) mode was developed for NaDCC (Sodium dichloroisocyanurate) quantification. This fast and low-cost method can be used to quantify NaDCC solid dosage forms using ATR-FTIR in absorbance mode in conjunction with partial least squares. A simple sampling procedure is included in the proposed experiment by just dissolving the samples in deionized water. An algorithm pipeline is also included for data cleaning, such as outlier removal, scatter correction, scaling, and mapping of the sample's spectrum to a NaDCC concentration. In addition, a simple model based on Beer's law was evaluated on a sub-range of $1220-1830\,\mathrm{cm}^{-1}$. Furthermore, a variable selection algorithm shows minimum excipient interference from the sample matrix in addition to visual analysis. A statistical analysis of the proposed method shows that it demonstrates a promising result with a regression coefficient of 0.996 ($R^2 = 0.996$) and recovery range of 95.5%–107%. As a result of the positive correlation of ATR-FTIR with NaDCC concentration, and in conjunction with the proposed method, this can serve as a clean, fast, affordable and eco-friendly method for pharmaceutical analysis.

Keywords: Machine learning · ATR-FTIR · Chemometric

1 Introduction

One third of people worldwide lack access to safe drinking water, with significant consequences for health [26]. Ensuring the availability of water is one of the United Nations Sustainable Development Goals [25]. Beyond systemic problems of service provision, water may be contaminated or temporarily restricted

© The Author(s) 2023
L. Longo and R. O'Reilly (Eds.): AICS 2022, CCIS 1662, pp. 106–120, 2023.
https://doi.org/10.1007/978-3-031-26438-2_9

during disasters, and providing emergency supplies and short-term purification treatments are essential. Water treatment and disinfection can be accomplished by methods including boiling, filtration, distillation and chlorination [24]. Chlorination is fast and effective, and can be delivered as soluble tablets or chemicals. Water disinfection tablets, with NaDCC (Sodium dichloroiso-cyanurate) as their main chemical component, have been shown to outperform iodine tablets for biocidal and cysticidal treatments [11,12].

Data-driven techniques for spectroscopy analysis are now common in chemometrics, and classical machine learning approaches are competitive with data-heavy neural network methods [19]. In this paper, we propose a new method which exploits machine learning and data analytics methods to quantify NaDCC, as a replacement for the slow and expensive laboratory techniques of titration [19]. Partial Least Squares regression (PLS) [10] is applied to the tablet formula solution spectra. The method is fast enough to perform during processing as part of a batch failure prevention test, and does not require significant additional expertise on behalf of the operators. Further, we show that its accuracy is within the same 90% to 110% recovery rate as the current method.

1.1 Pharmaceutical Background

Tablets are the most common solid dosage form for pharmaceutical products due to inexpensive manufacturing, packaging, transportation costs and popularity [13]. Active pharmaceutical ingredient (API) quantification in quality control is an integrated part of the tablet manufacturing life-cycle [6]. Aside from the formula, the concentration of each component in a solid dosage form is determined by other process factors, including powder flow, particle size distribution, dosing depth and turret speed. Blend uniformity is an important quality test that checks for uniform distribution of content in the mixture which has a direct impact on the average quantity of API per tablet [2].

HPLC and capillary electrophoresis [9] are extensively used and considered as standard methods for quantitative testing of different pharmaceutical formulas. These methods require significant amounts of sample preparation and analysis time, in addition to the very high cost of these instruments. Therefore, the need for quick, cost-effective, and easy-to-use technologies for quality control, such as FTIR spectroscopy arises [15]. ATR-FTIR (Attenuated Total Reflection-Fourier Transform Infrared) can be used instead of standard methods to assay API. Different materials absorb infrared light in different patterns depending on whether they have a covalent bond that vibrates at a specific frequency [17]. This enables it to detect molecular vibrations and identify specific chemicals [1]. Infrared spectroscopy has been widely investigated for both qualitative and quantitative analysis of pharmaceutical analysis [8,15].

In a batch manufacturing process, to evaluate the concentration of NaDCC in a batch, a sample of tablets is normally taken during production and after it is completed. In case the concentration is not within specification, three more samples will be collected, and if the concentration is still out of specification, the batch will be rejected. The time required to perform any of these assays is around 30 min, which makes it impossible to use them for in-process quality

control. Since the tablet formulation is the same for each product, and sometimes batch failures occur, we should investigate the production process to determine the root cause. The two main steps in the process are blending and compression, and since the blending configuration does not change during a batch, compression is the likely cause of any batch failure. The rotary tablet press machines used for compression must be clean and undergo regular maintenance. There may be times when a rotary tablet press is used for another product, in which case the configuration will change according to the new product's requirements. In addition, depending on the product, there is also quality control during the manufacturing process. A change of configuration will be made to the rotary tablet press if the results of these quality tests fail to meet specifications. These in-process quality checks are entangled with each other. Turret speed has a positive association with weight variation and a negative correlation with die filling, resulting in weight and hardness that are both out of specification. Also, hardness has a negative correlation with paddle speed and positive correlation with die depth [21]. In order to bring all of these in-process tests into specification the operator might need to configure the rotary tablet press so that some of these metrics are at their specification boundary. In addition, some of these in-process quality checks directly influence the concentration of active biocides such as NaDCC. Weight and NaDCC concentration are positively correlated, for instance. For all of these reasons, it would be beneficial to find an alternative approach to NaDCC quantification as part of the manufacturing process control.

The ATR-FTIR spectrum of solution of water purification tablets and chemometrics techniques were used to explore NaDCC quantification in this work. In simple terms, samples are prepared by dissolving solid dosage forms in deionized water, spectrum recordings are made from that solution, and the concentration is quantified based on the pipeline proposed for the prediction algorithm. The proposed economical and quick approach has potential as an alternative to the current techniques which are slow, require detailed method development and tedious sample preparation techniques.

2 Experiment

Medentech, Wexford, Ireland, supplied three excipients and one API: sodium bicarbonate, sodium carbonate and adipic acid as excipients, and NaDCC as an active biocide. For a successful measurement, various factors such as humidity, content distribution uniformity and temperature must be taken into account. To circumvent the difficulties noted in Sect. 1.1, samples were dissolved in deionized water obtained from an Elix® Advantage 3 Water Purication System. The deionized water used had a conductivity lower than 0.2 µS/cm at 25 C, the resistance was greater than 5 mOhm-cm, and the organic carbon content was less than 30ppb. After each sample recording, ESEPT® alcohol-based Isopropanol 70 v/v was used to clean the surface of the ATR accessory.

A basic sampling approach was employed to establish a quick and easy procedure which can be used for in-process quality control. Samples were prepared in disposable plastic containers. The scale was calibrated to zero while the sampling

container was on the scale. Each component was removed from its bag and placed in the container with a clean spatula until the needed amount was attained. All of the components were weighed using the same approach. Samples were then gently placed into a 500 ml beaker containing 200 ml of deionized water. The samples were then dissolved thoroughly in deionized water in order to form a homogeneous solution. This was achieved by sonicating them for 2 to 10 min, depending on the sample. For both formulas, the above steps were followed to prepare the sample. Samples begin with zero concentration of NaDCC, and the quantity of excipients was decreased while gradually increasing the amount of NaDCC, so that the overall amount of the blend (combination of excipients plus NaDCC) remained constant. Following sonication and homogenization, one drop of solution was taken with a pipette for examination. Next, the same sample was diluted with 100 ml of deionized water and was processed as an independent sample. The dilution process was repeated three times, and each diluted sample was considered as a separate sample. The beakers and tools were thoroughly cleaned after each sample and re-used only after they were completely dry. Twenty samples, each diluted four times, were collected.

In this experiment, a Compact Alpha P FT-IR Spectrometer (Billerica, Massachusetts, United States), equipped with a diode laser with spectral stability and high wavenumber accuracy was used. All measurements were taken by a high-performance Platinum-ATR accessory featuring a monolithic trapezoid shape diamond crystal. Three spectra were acquired for each sample, and each spectrum was scanned 24 times. The spectral range was $4000-400\,cm^{-1}$ and the Spectral Resolution was $2\,cm^{-1}$. Each spectrum gives 1776 data points on 2.04 wave-number intervals. Dissolution of the samples was performed using an ultrasonic bath, Decon FS200. Each sample is sonicated for 2–10 min to produce a homogenous solution. A magnetic stirrer was used for between three and twelve minutes, depending on adipic acid concentration.

Blends were created based on the formula for water purification tablets produced by Medentech. These blends were prepared on a small scale, so each sample weighed approximately 20 g. There are three excipients examined within each blend (sodium bicarbonate, sodium carbonate, and adipic acid) and one active pharmaceutical ingredient (NaDCC). Having performed this step, the concentration of NaDCC in each sample was established. The next step was to dissolve the blends in deionized water. Each blend (20 g) was dissolved in 200 ml deionized water in a 500 ml beaker. A homogeneous solution was obtained by sonicating the samples for 2–10 min. Additionally, each sample was diluted three times with 100 ml of deionized water each time in order to collect more data. The spectrometer prism was cleaned with Isopropanol alcohol (IPA) after recording each sample. The sampling surface of the spectrometer needs 15 s to completely dry out. Before sampling, the spectrometer was set at a resolution of $4\,cm^{-1}$ and a range between $4000\,cm^{-1}$ and $400\,cm^{-1}$. A disposable plastic pipette was used to place one drop of solution on the prism of the spectrometer to record sample spectra. Three scans were conducted on each sample. The recording of any sample was preceded by a background scan. There were 24 scans each for the background spectrum and the sample spectrum.

3 Results and Discussions

Excipients are added to an active pharmaceutical component (NaDCC) in Medentech's water purification solid dosage forms for a variety of purposes. The basic goal is to enhance the formulation's volume, make packaging and transportation easier, and impart desirable qualities [23]. The proportion of each excipient varies depending on the product's use case. In Fig. 1, four different sample spectra are shown: water, pure NaDCC, a normal sample based on Medentech product formulation, and a sample containing only excipients. In the range 1000 cm^{-1} to 2000 cm^{-1}, both NaDCC and the excipients show two additional peaks. In particular, NaDCC shows a unique peak at approximately 1250 cm^{-1}. Therefore, the following methods focus on that 1000−2000 cm^{-1} range.

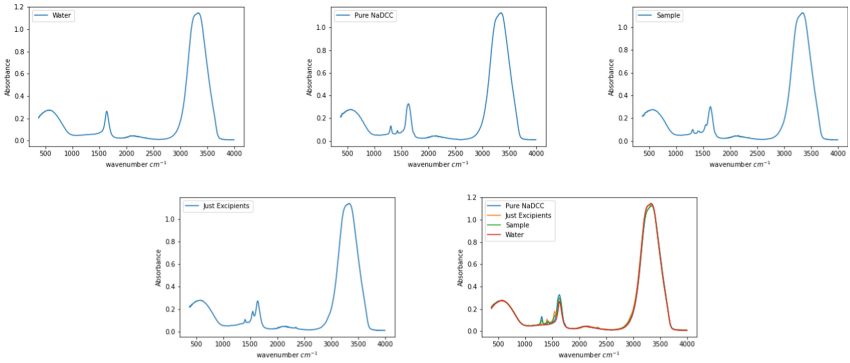


Fig. 1. Different concentrations of NaDCC and excipients in deionized water.

3.1 ATR FTIR Region Selection for Calibration Models

Figure 2 illustrates various sample solution spectra with different NaDCC concentrations. We see two peaks at approximately 1400 and 1550 where the absorbance seems to be inversely proportional to the NaDCC concentration. This is because the absorbance of the excipients at these wavenumbers is higher than for NaDCC (as seen in Fig. 1). Peak intensity of the spectrum (A) depends on molar absorptivity (ϵ), path length (b) and concentration (c) ($A = \epsilon bc$). If two substances have absorption at the same wavenumber and they are both present in a sample the response of the instrument depends on the concentration of each substance and their molar absorptivity. Since the amount of powder dissolved in the deionized water was held constant at 20g, when the concentration of NaDCC is increased, the concentration of excipients was decreased, and so the overall absorbance of the sample decreases at those wavenumbers. We therefore focus on the outer two peaks, where the absorbance of NaDCC is higher than for the

excipients, again as seen in Fig. 1. For these two peaks, absorbance is approximately linear with respect to concentration, and so Beer's law [22] can be applied. Figure 2 illustrates where the height fluctuates with NaDCC and there are four bands evident. This is confirmed through formal analysis, both univariate, in the spectrum range 1220–1830, and multivariate, in range 400–4000.

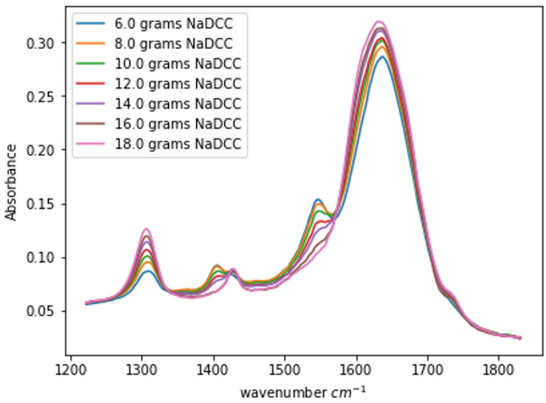


Fig. 2. ATR FT-IR spectrum of NaDCC concentrations in 200 ml of deionized water.

The response of a calibrated univariate Beer's law model in the spectra in the range 1220–1830, according to peak height - with or without baseline modifications - is shown in the Table 1. The average recovery column of this summary shows the average recovery of at least 5 samples with the same concentration of NaDCC. In the table, we can see that the baseline correction [14] slightly enhances the accuracy and precision of Beer's law calibrated models. The results confirmed the visualization assessment shown in Fig. 2, which shows a high correlation between wavenumbers in the range 1220–1830 and the concentration of the target analyte. Wavenumber 1281 has the highest correlation coefficient (0.9971) and and an average recovery rate of 100.93.

3.2 PLS Calibration Model

Partial least squares regression plays an important role in chemistry [10] and high-dimensional collinear data processing. This technique resolves the multicollinearity problem associated with most spectroscopy data sets by mapping the acquired data into a set of latent variables [4] of much smaller size. A data preprocessing step is typically incorporated into statistical analysis and modeling along with the main prediction model, so preprocessing steps are vital to getting the most out of machine learning algorithms. Some algorithms might require all of these preprocessing steps, while others might require only a subset.

Table 1. The peak height selection of FTIR area and wavenumber.

Wavenumber	Baseline criteria	Regression	Corr. coefficient	Recovery Rate
1283	None	4.210x + −0.257	0.9959	100.67
1285	None	3.595x + −0.22	0.9962	100.93
1523	None	−2.338x + 0.3	0.9964	99.58
1534	None	−1.492x + 0.233	0.996	100.45
1536	None	−1.403x + 0.226	0.9961	100.76
1279	400–4000	5.2352x + 0.1171	0.9967	101.34
1281	400–4000	4.4754x + 0.0998	0.9971	100.93
1283	400–4000	3.8356x + 0.0854	0.9965	100.82
1536	400–4000	−1.4613x + 0.1403	0.9967	100.62
1538	400–4000	−1.3854x + 0.1394	0.9963	100.97

We use an algorithm pipline of outlier detection, smoothing, scatter correction, variable selection, and PLS, to produce the NaDCC quantification from the input spectrum. Standard procedures for developing algorithms in machine learning and data analysis are used in the development of this pipeline. Three main steps comprise the general pipeline: data collection, preprocessing, and model prediction. An ATR-FTIR sample scan was done as a first step, based on the prepared sampling method, to gather raw data. Following that, general and specific data preprocessing steps are undertaken such as normalization, artifact removal, smoothing, and variable selection. In this study, PLS was calibrated using k-fold cross validation, and evaluated using R-square for unseen samples.

When a data point does not fit the general trend, it is usually considered an outlier. A model is not able to explain outliers well because they are associated with large errors in the cost function. There are several causes of outliers, including measurement error, sampling error, inaccurate recording, or incorrect assumptions about the distribution. In the outlier identification approach, the Q-residual, Hotelling T-Square, which is capable of reducing the computation time without compromising accuracy [16], and a 95% confidence interval were employed. According to Fig. 3, 19 cases were detected as outliers. Additionally, an approach can be used for eliminating outliers simply by examining samples visually. Figure 4 shows the effect of one poor quality scan (with three scans per sample), where the curve can be seen to deviate significantly from the curves for the other sample scans. Any given model will contain a data point that has a high Q-residual in comparison to the corresponding residuals of other data points, so there will always be some data points with a large residual error. In this study, model calibration by removing outliers based on the output of Q-residual and Hotelling T-Square approaches was applied to avoid the proposed model under-performing.

Signal smoothing is just as critical to the pre-processing of spectral data as removing outliers from the data points. The random noise in the data can be

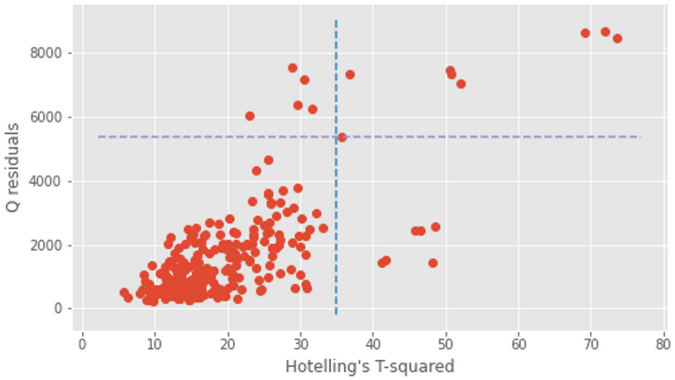


Fig. 3. Q-residual and Hotelling T-Square metrics Outlier Detection

reduced using a smoothing technique. Many methods can be employed to accomplish this task, such as Savitzky-Golay, and Fourier spectral smoothing [27]. In this model, for each point in the sample, a neighbourhood of points were selected, which is called the window size, and then a polynomial model is fitted to the selected data points in the window. This data point is replaced with the corresponding value of the fitted curve at that point in order to provide the smooth version of the data point. The Savitzky-Golay results in the suggested pipeline design were calculated using a window size of 11 and a polynomial degree of 3 (Fig. 5).

Specular reflections and diffuse reflections constitute a spectrum. Due to the sample's chemical composition, different wavelengths of incident light are absorbed differently by the sample, resulting in different spectral shapes. Additionally, particle sizes and path lengths also affect spectra. Scattering can be used to eliminate errors caused by sample geometry and morphology which have no connection to chemical composition. In essence, removing all these undesirable effects before computing the quantity of interest produces a better model. There are two tools that can be used in spectroscopy to correct scatter data-the standard normal variations (SNVs) and the multiplicative scatter correction (MSCs) [7]. The particle size and path length effects are expected to have a zero mean normal distribution in each sample, and these scatters should be reduced significantly by averaging all samples. An average spectrum is calculated from all the samples and a linear regression model is fitted to the calculated spectrum as an independent variable and to each sample as a dependent variable. Equation 1 illustrates the general procedure for the MSC.

$$X_m = \frac{1}{N}\sum_{i=1}^{N} X_i \quad X_i = b_i + m_i \times X_m \Rightarrow X_i^{msc} = \frac{X_i - b_i}{m_i} \tag{1}$$

In the SNV method, used in this study (Fig. 5), there is no reference to regress the input spectrum against. We use $X_i^{snv} = \frac{X_i - \hat{X}_i}{\sigma_i}$.

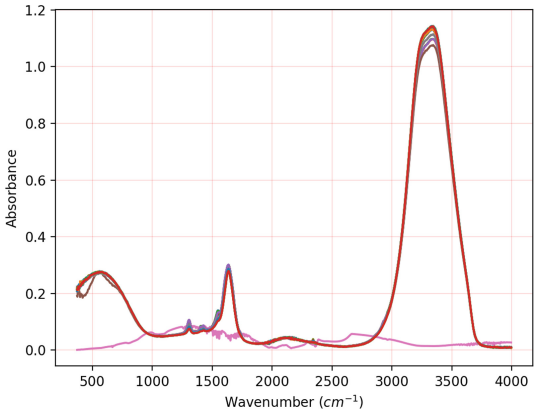


Fig. 4. Examples of a misaligned sample

Pharmaceutical laboratories heavily rely on instruments that generate large amounts of data. It is fairly common for a laboratory instrument to generate data that has thousands of variables; for example a typical FTIR instrument records absorbance at more than 10000 frequencies. However the full amount of this data is not useful in many of scenarios and normally there is considerable redundancy and correlation among these variables. Since high dimensional data leads to problems related to the curse of dimensionality in machine learning, extracting and compressing these variables in such a way that keeps essential information is vital. This compression or extraction may be achieved by combining different variables to get a more informative variable (such as Principal component analysis-PCA) or by selecting a variable from a set of variables that provide more information for the task in hand.

Due to the simplicity of the forward variable selection algorithm, it is applied to this problem in order to ensure that important wavelengths are separated from less informative wavelengths within spectral measurements [18]. The wavelength bands containing most of the signal related to the analyte can often be hard to predict in advance, especially in visible and infrared spectroscopy. A measurement of all bands that the instrument is capable of will be made in the first step, followed by a determination of vital bands. In other words, the wavelength bands will make better-quality models stand out.

Preprocessing discards one wavelength at a time. An entire spectrum calibration model will be created, then the wavelength associated with the regression coefficient with the smallest absolute value will be eliminated. By calculating the model's mean square error in each iteration, the performance will be evaluated. A given number of wavelengths will be discarded in conjunction with the minimum model mean square error. A total of 1292 wavelengths were discarded and the optimised MSE was 0.006929.

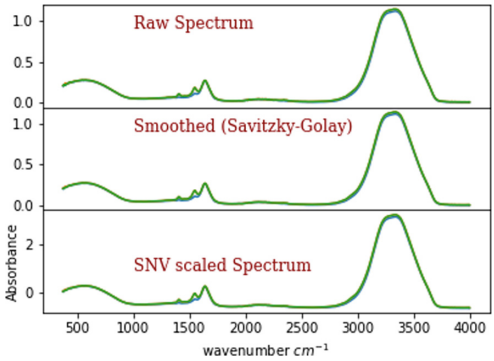


Fig. 5. Spectrum smoothing and scatter correction. The top graph displays raw spectrum in absorbance mode with no processing; the middle graph is obtained by applying a savitzky golay filter with window size 11 and degree 3; the bottom graph illustrates the spectrum after applying the standard normal variations scatter correction.

The estimation part of the pipeline consists of partial least square regression (PLSR) as its core algorithm. The concentration is predicted by this algorithm after the preprocessing stage, where the spectrum is mapped to concentration. In partial least square regression, multiple linear regression is performed, which builds a linear model, $Y = XB + E$, which maps latent variables (LV) onto dependent variables. The method is designed to maximize correlation between the selected LVs and the target variable. The reason that PLSR is superior to PCR (principal component regression) is that it simultaneously extracts variability from input data (X) and correlates it with target data (Y). In PCR, the LVs of PCA are used to account for variation in independent variables, but they may not affect the dependent variable directly.

In spectroscopy analyses, three variables are involved: X, Y, and E. X represents spectra, Y represents quantities or quantity sets, and E represents errors. In mathematical terms, PLSR can be considered as an optimization problem with the objective: $arg\max_{w_i} cov(XW, Y) \quad i = 1, \cdots, A$. The analysis of the collected data was based on the PLSR model. Since 80 samples were gathered and each sample was scanned three times, a total of 240 samples were collected. Data was split as a typical procedure in machine learning and chemometric analysis, ten percent for testing, and ninety percent for calibration. During variable selection 1776 variables were reduced to 484, which were used as input for model calibration, reserving 24 samples as the test set. Variables were selected based on the magnitude of mean square error produced by the PLSR model on the calibrated dataset. The optimal number of variables was associated with the minimum error of the model. The model was calibrated based on 10-fold cross-validation in the selection of variables as well as in the training process for predicting the interest target. In addition, the optimal number of PLS components for the PLS model had to be estimated. The search for a model includes all potential component combinations between 1 and 30. On the other hand,

according to Fig. 6, PLS models with 15 components produce the smallest average square error, which aligns with the optimal number of components in the variable selection stage.

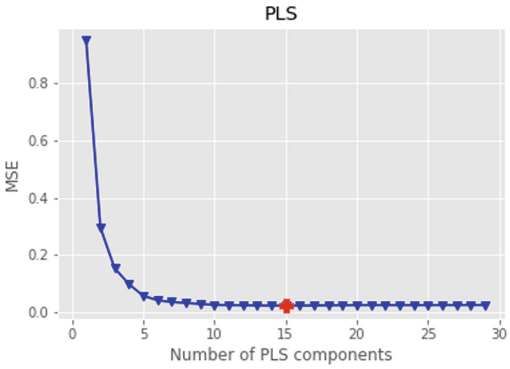


Fig. 6. Optimal number of component for PLSR

The minimum error of the calibration model, illustrated in Fig. 6, is responsible for the good dispersion of predicted values of NaDCC concentration around the regression line. In addition to determining how well the calibration model fits, a second factor to consider is the square of correlation coefficient (R^2). This is a measure of how well the independent variables can explain the variation in dependent variables. The R^2 value, 0.9961, is shown in Fig. 7. The value is over 0.99, which is representative of a high degree of linear correlation between the predicted and the ground truth values.

NaDCC Quantification Result. This is the last step of the algorithm pipeline, which evaluates the algorithm's performance capabilities and statistical analysis. Seven test groups are used to perform this evaluation. Each group consists of several samples with the same NaDCC concentrations, but each group's number of samples varies. Occasionally, after dilution of a sample, the concentration of API in two different samples was equal. The purpose of the diluting procedure was to create additional data. A summary of the results of our model is presented in Table 2. Within the 7 test groups, the recovery average ranged from 95.46% to 107%, which is in complete agreement with the baseline (titration) that we are comparing to which has a target recovery range of 90% to 110%. Another important evaluation metric in chemometric analysis is the limit of detection (LoD) and limit of quantification (LoQ). LoD is the least quantity of analyte that can be consistently distinguished from zero concentration, whereas LoQ is the smallest value of analyte that can be quantified [5,20].

The LoD in relation to partial least square regression has been calculated using equation 3 from Franco et al [3]. In their proposed LoD formula, excipient-containing samples will be treated as samples with zero concentration. We will

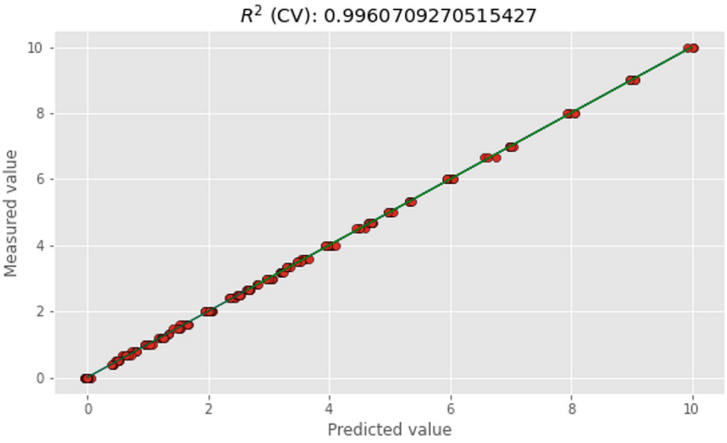


Fig. 7. Test Coefficient of determination ($R^2 - Score$)

calculate LoD using $LoD = (t_{\alpha,\nu} + t_{\beta,\nu}) \times \sqrt{var(y_0)}$, in which y_0 represents a sample with a concentration of zero, and $t_{\alpha,\nu}$ and $t_{\beta,\nu}$ represent the parameters of a t distribution that has ν degrees of freedom. The limit of detection has been calculated at 0.0849 mg/ml, while the limit of quantification (LoQ) has been calculated at 0.283 mg/ml.

Table 2. Samples

Sample	NaDCC measured	NaDCC predicted	Recovery average	Standard deviation
1	1.000	0.955	95.462%	0.006
2	1.500	1.616	107%	0.068
3	1.600	1.536	96%	0.095
4	2.000	1.985	99.237%	0.015
5	3.200	3.360	105.012%	0.066
6	5.000	4.909	98.188%	0.057
7	6.000	6.245	104.076%	0.028

4 Conclusion

This study proposes applying data analysis techniques directly to the FTIR spectrum of chemical compounds (ATR-FTIR) to quantify APIs of interest. The method is to eliminate the complicated traditional, time-consuming titration methods, simplify sampling procedures, and expedite result extraction for in-process quality control. The method's low cost and the elimination of toxic chemicals from titration methods makes it environmentally friendly. By simply

dissolving NaDCC tablets in deionized water and using it as an ATR-FTIR sample, the proposed method successfully quantifies NaDCC concentrations. According to an evaluation of the proposed pipeline with $R^2 = 0.996$ and recovery range of 95.5%–107%, which completely aligned with the recovery range required in the case study. Additionally, the process can be completed in less than 3 min, making it suitable for use as an in-process quality control method. The technique could potentially replace the existing labor-intensive and time-consuming titration technique for analysis of NaDCC concentrations.

Acknowledgements and Data Availability. This publication has emanated from research supported in part by Science Foundation Ireland under Grant number 16/RC/3918 which is co-funded under the European Regional Development Fund. The datasets generated during the current study are available in google drive: NaDCC.db.

References

1. Abd El-Rahman, M.K., Eid, S.M., Elghobashy, M.R., Kelani, K.M.: Inline potentiometric monitoring of butyrylcholinesterase activity based on metabolism of bambuterol at the point of care. Sens. Actuators B: Chem. **285** (2019)
2. Akseli, I., et al.: A practical framework toward prediction of breaking force and disintegration of tablet formulations using machine learning tools. J. Pharm. Sci. **106**(1), 234–247 (2017)
3. Allegrini, F., Olivieri, A.C.: IUPAC-consistent approach to the limit of detection in partial least-squares calibration. Anal. Chem. **86**(15), 7858–7866 (2014)
4. Boulesteix, A.L., Strimmer, K.: Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief. Bioinform. **8**(1), 32–44 (2006)
5. Currie, L.A.: Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC recommendations 1995). Pure Appl. Chem. **67**(10), 1699–1723 (1995)
6. Davidson, I.E.: 11 - setting up specifications. In: Ahuja, S., Scypinski, S. (eds.) Handbook of Modern Pharmaceutical Analysis, Separation Science and Technology, vol. 3, pp. 387–413. Academic Press (2001)
7. Dhanoa, M., Lister, S., Sanderson, R., Barnes, R.: The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. J. Near Infrared Spectrosc. **2**(1), 43–47 (1994)
8. Eid, S.M., Soliman, S.S., Elghobashy, M.R., Abdalla, O.M.: ATR-FTIR coupled with chemometrics for quantification of vildagliptin and metformin in pharmaceutical combinations having diverged concentration ranges. Vib. Spectrosc. **106**, 102995 (2020)
9. ElBagary, R.I., Azzazy, H.M., ElKady, E.F., Farouk, F.: Simultaneous determination of metformin, vildagliptin, and 3-amino-1-adamantanol in human plasma: application to pharmacokinetic studies. J. Liquid Chromatogr. Relat. Technol. **39**(4), 195–202 (2016)
10. Geladi, P., Kowalski, B.R.: Partial least-squares regression: a tutorial. Anal. Chim. Acta **185**, 1–17 (1986)
11. Gerba, C.P., Johnson, D.C., Hasan, M.N.: Efficacy of iodine water purification tablets against cryptosporidium oocysts and giardia cysts. Wilderness Environ. Med. **8**(2), 96–100 (1997)

12. Kgabi, N., Mashauri, D., Hamatui, N.: Utilisation of water purification "tablets" at household level in Namibia and Tanzania. Open J. Appl. Sci. **4**, 560–566 (2014)
13. Kottke, M., Rudnic, E.: Tablet dosage forms. In: Banker, Rhodes (eds.) Modern Pharmaceutics, pp. 458–532. CRC Press (2002)
14. Liland, K.H., Almøy, T., Mevik, B.H.: Optimal choice of baseline correction for multivariate calibration of spectra. Appl. Spectrosc. **64**(9), 1007–1016 (2010)
15. Mallah, M.A., Sherazi, S.T.H., Bhanger, M.I., Mahesar, S.A., Bajeer, M.A.: A rapid Fourier-transform infrared (FTIR) spectroscopic method for direct quantification of paracetamol content in solid pharmaceutical formulations. Spectrochim. Acta Part A Mol. Biomol. Spectrosc. **141**, 64–70 (2015)
16. Mashuri, M., Ahsan, M., Lee, M.H., Prastyo, D.D.: Wibawati: PCA-based hotelling's T2 chart with fast minimum covariance determinant (FMCD) estimator and kernel density estimation (KDE) for network intrusion detection. Comput. Ind. Eng. **158**, 107447 (2021)
17. Mayerhöfer, T.G., Popp, J.: Beer's law - why absorbance depends (almost) linearly on concentration. ChemPhysChem **20**(4), 511–515 (2019)
18. Mehmood, T., Liland, K.H., Snipen, L., Sæbø, S.: A review of variable selection methods in partial least squares regression. Chemom. Intell. Lab. Syst. **118**, 62–69 (2012)
19. O'Connell, M.L., Howley, T., Ryder, A.G., Leger, M.N., Madden, M.G.: Classification of a target analyte in solid mixtures using principal component analysis, support vector machines, and Raman spectroscopy. In: Opto-Ireland 2005: Optical Sensing and Spectroscopy, vol. 5826, pp. 340–350. SPIE (2005)
20. Saadati, N., Abdullah, M.P., Zakaria, Z., Sany, S.B.T., Rezayi, M., Hassonizadeh, H.: Limit of detection and limit of quantification development procedures for organochlorine pesticides analysis in water and sediment matrices. Chem. Cent. J. **7**(1), 63 (2013)
21. Schomberg, A.K., Kwade, A., Finke, J.H.: The challenge of die filling in rotary presses-a systematic study of material properties and process parameters. Pharmaceutics **12**(3) (2020)
22. Swinehart, D.F.: The beer-lambert law. J. Chem. Educ. **39**(7) (1962)
23. Tekade, R.: Basic fundamentals of Drug Delivery. Academic Press, an Imprint of Elsevier, London, San Diego (2019)
24. Torricelli, A.: Drinking water purification. Adv. Chem. **21**, 453–465 (1959)
25. United Nations Department of Economic and Social Affairs: Goal 6: Ensure availability and sustainable management of water and sanitation for all
26. World Health Organization: 1 in 3 people globally do not have access to safe drinking water. https://tinyurl.com/3dkfupct
27. Zhao, A.X., Tang, X.J., Zhang, Z.H., Liu, J.H.: The parameters optimization selection of Savitzky-Golay filter and its application in smoothing pretreatment for FTIR spectra. In: 2014 9th IEEE Conference on Industrial Electronics and Applications, pp. 516–521. IEEE (2014)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

