Implication of phenol red in quantification of cultured cancerous cells using near-infrared spectroscopy and aquaphotomics

Raypah, E. Muna ; Muncan, Jelena ; Sudik, Suhainah ; Omar, Fairuz Ahmad ; Mail, Hafiz Mohd ; Tsenkova, Roumiana ; Seeni, Azman

(Citation)
Chemometrics and Intelligent Laboratory Systems,230:104669

(Issue Date)
2022-09-21

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/0100488647

Implication of Phenol Red in Quantification of Cultured Cancerous Cells using Near-Infrared Spectroscopy and Aquaphotomics

Muna E. Raypah [1*], Jelena Muncan [2**], Suhainah Sudik [1], Ahmad Fairuz Omar [1***], Mohd Hafiz Mail [3], Roumiana Tsenkova[2], and Azman Seeni [4]

[1] School of Physics, Universiti Sains Malaysia, 11800, Pulau Penang, Malaysia
[2] Aquaphotomics Research Department, Faculty of Agriculture, Kobe University, Kobe, Japan
[3] Malaysian Institute of Pharmaceuticals and Nutraceuticals, National Institute of Biotechnology Malaysia, Ministry of Energy, Science, Technology, Environment and Climate Change, 11700 Penang, Malaysia
[4] Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200, Pulau Penang, Malaysia
*Corresponding authors' e-mails: fairuz_omar@usm.my, muna_ezzi@usm.my, , and jmuncan@people.kobe-u.ac.jp

Abstract

This study aims to evaluate the ability of absorbance spectra in the near-infrared (NIR) region to predict the number of cells for different cell lines. The cancerous cell lines were human cervix adenocarcinoma (HeLa) and human prostate carcinoma (DU145), and L929 was a normal mouse skin fibroblast. The number of cells varied from 50,000 to 275,000 with an interval of 25,000 for each cell line. Vis-NIR absorbance spectra (400-1100 nm) at each number of cells (50,000-275,000) for L929, DU145, and HeLa cell lines cultured in media with and without phenol red were recorded. Multiple linear regression (MLR) and partial least squares regression (PLSR) models were developed in the NIR region (680-1050 nm) to quantify the number of cells for the three cell lines. The outcomes showed that the quantification analysis of the number of cells using MLR and PLSR models produced high prediction accuracy with $R^2 \geq 93\%$. The best results were obtained using PLSR for the preprocessed spectra using an orthogonal signal correction method. It was found that the presence of phenol red in the culture medium improved the prediction accuracy in the case of HeLa (SECV = 271 cells) and DU145 (SECV = 250 cells) cell lines, but the accuracy was higher when phenol red was not present for the L929 cell line (SECV = 24 cells). In addition, the existence of phenol red boosted the accuracy when the global PLSR model was built, irrespective of the cell type (SECV = 7,754 cells). The effect of phenol red was explained in the terms of its impact on the water molecular structure of the cells' culture medium which influences the light scattering.

Keywords: Cancerous Cells; Phenol Red; Near-infrared Spectroscopy; Quantification; MLR; PLSR.

1. Introduction

Cancer is a complex disease, typified by reprogrammed signaling, cellular physiological alterations, and uninhibited cell growth. A characteristic trait of cancer cells is their metabolic reprogramming, which facilitates fast cellular reproduction, migration, and alteration of their microenvironment, ultimately enabling metastasis [1]. Early detection and early-stage surgical and chemotherapeutic interventions and therapies help inhibit cancer growth. Nonetheless, cancer becomes untreatable or terminal when it metastasizes. The gradual growth of a tumor allows for the early detection of a malignancy (or pre-malignancy) which lowers the prevalence of late-stage cancer [2]. The early detection of tumors is crucial for effective cancer treatment and for elucidating tumorigenesis. Expanding the range of recognized reliable biomarkers may aid the discrimination of benign cancers from malignant ones by clinicians, thus decreasing the incidence of needless biopsies, and enhancing the effectiveness of cancer treatments.

The recent advancement in point-of-care diagnostics and precision medicine has prompted the demand for inexpensive and accurate cell counting technologies [3]. The proliferation of tumor cells is determined primarily by three key parameters: the progression period for reproducing cells, a part of proliferating cells, and the quantity of unprompted cell damage [4]. Growth rates of solid tumors can be determined without difficulty along the linear portion of the Gompertzian growth curve. Visible and detectable tumors typically hold between 10^6 to 10^9 cells [5, 6]. The current detection threshold for solid tumors proliferating as a single mass is in the region of 10^9 cells ($1\ \mathrm{g} = 1\ cm^3$) [5]. Numerous cases of different solid tumors have shown linearity of growth on a logarithmic scale throughout the early detection stage.

Globally, treatment centers and hospitals extensively utilize cell counting techniques for the determination of a patient's health [3]. Clinically, a tumor with a volume of 1 cm^3 ($\approx 10^9$ cells)

is considered rather small, and at this size, the tumor may induce the initial symptoms and can be perceptible on palpation, or with the use of laboratory diagnostic examinations. The most frequent error is to consider a small tumor as an ‘‘early’’ tumor. At the stage when it becomes detectable, a malignant tumor has existed for a minimum of half of its lifespan [4]. The significance of early tumor detection for the diagnosis of the patient has been overly estimated since tumors cannot be clinically detected early. Cancer may have spread long before the primary tumor can be diagnosed. The precise determination of the number of cells in culture is crucial for experimental reproducibility and standardization [7]. Cell counts are vital for the evaluation of cell health and proliferation rate, assessment of immortalization or alteration, seeding cells for successive experiments, transfection or infection, and preparation for cell-based assays. The cell counts must be precise, reliable, and rapid, especially for the quantitative determination of cellular responses [8]. There is also an immense need for the precise quantification of biomarkers for cancers [9], particularly at the early recognition of a miniature volume of tumors [10].

Quantitative techniques are advantageous due to the insight they offer into understanding the biology/physiology of tumor cells beyond diagnosis. The method that is most commonly used for direct cell counting is trypan blue staining using a hemacytometer. This approach allows counting the number of cells manually under a microscope. However, the hemacytometer-based cell enumeration can be very biased, subject to errors due to device misuse, tiresome, and time-consuming. Coulter counter (automated counter) was the first electronic cell counter which was widely accepted as an alternative to manual counting. Automated cell counters based on the Coulter counter principle, optical flow cytometry, and image-based cytometry have been established and commercially available. These devices can detect cells using integrated software to describe cells in terms of their fluorescent intensity and size in a few seconds. It was reported a

high precision and accuracy were obtained in the case of using the optical flow cytometer in comparison to manual and Coulter counting methods [11]. Currently, in research laboratories, images of cells obtained using microscopy are also commonly examined and quantified using open-source software [12]. Cell profiler, ImageJ, and other such open-source software are routinely employed by cell biologists for cell enumeration from images [13-15]. This can be achieved by performing object identification through segmentation, thresholding, recognition, and division of clumped cells and other processing steps. At present, most of the microfluidics-based cell enumerating systems available in the market utilize a microfluidic chip or cartridge combined with a traditional cytometry system [16]. Such designs offer ease and flexibility since they include a fluidics system with microchannels for sample flow. However, the microfluidic devices still need more improvement in sample preparation and detection techniques within a lab-on-a-chip microfluidic platform [3]. Therefore, there is increasing demand for an objective cell quantification technique that produces an accurate and reliable number of cells.

The quantification of predictive biomarkers for cancer is significant, particularly at the early stage of cancer when the number of cells is small. In addition, the quantification of cells is essentially significant in both fields of pathology and clinical studies [17] and is considered a vital approach to elucidating the cellular composition, growth, development of diseases, and aging. Recently, the optical spectroscopy technique has unveiled its potential undamaging capability to exhibit distinct spectral attributes that discriminate between tumors and normal cells. The spectroscopy techniques are rapid, simple to use, environmental-friendly, cost-effective, and non-destructive. Thus far, spectroscopy techniques have been extensively utilized for the diagnosis of a range of cancer types such as lung, oral, leukemia, breast, and prostate cancer [18-24]. However, there is inadequate data on the optical spectroscopy properties of cancer cells, particularly for the

quantitative assessment [25]. The potential to attain quantitative characteristics of cells directly from the living system in situ is important. There is a need to comprehensively characterize and establish spectroscopy signatures of cancer at the cell level to better understand the spectral responses. So far, information on cancer cells using the NIR spectroscopy technique is not extensively available, especially for the quantitative assessment of the cultured cells. The supposed "optical window" that occurs in the NIR region 650-1100 nm is appropriate for most of the non-invasive measurements of biological systems [26]. This region is also referred to as the "therapeutic window" for the lower and determinate light absorbance and the abundance of information.

The combination of analytical tools has been proven to be of immense value in a variety of bio-scientific applications. Spectroscopy coupled with chemometrics can offer a rapid method for improving the performance of the cell culture [18] and quantifying the number of cancer cells. Partial least squares regression (PLSR) [27, 28] is one of the most commonly used regression methods for quantification of multivariate spectral data together with multiple linear regression (MLR) and principal component regression (PCR) [29]. It is important to point out that the MLR model is a causative model with clear physical meaning and more explanatory coefficients compared to the coefficients of PLSR. On the other hand, in quantitative studies, the MLR model sometimes cannot provide predictive models which can be considered its main limitation [30]. PLSR is a good method of choice for modelling datasets where the number of variables (wavelengths) surpasses the number of observations (spectra of the samples) and where the predictors are highly correlated [27, 31]; this is considered the main advantage of the PLSR over MLR. In addition, the PLSR model with latent variables that extract the patterns characteristic of descriptor variables provides information about the predictive power of dependent variables [30].

Even though many researchers have worked on cancer cells, very few researchers have reported the quantification of the cancerous cells. Up to now, the spectroscopic analysis combined with chemometrics for quantifying the number of cancerous cultured cells in the NIR region with aid of phenol red has not been reported. This work is an extension of our previous study [25] with an intention to examine the implication of phenol red (a commonly pH indicator in culture media) towards the quantitative analysis of the cancerous cultured cells using the absorbance at the NIR region (700-1100 nm). This research aims to evaluate the accuracy of NIR spectroscopy and aquaphotomics to quantify the number of cultured cancerous cells by developing two regression models. The effect of phenol red was explained in the terms of its impact on the water molecular structure of the cells' culture medium which influences the light scattering. The cancerous cells utilized were cervical (HeLa) and prostate (DU145), while the mouse skin fibroblast (L929) cell line was used as a normal cell. Direct comparisons of the growth of the three cell lines were performed in the culture medium in the presence and devoid of phenol red dye. The MLR was used for a cross-calibration purpose and PLSR was used as another method for quantification of cells using the NIR region from 680 nm to 1050 nm. A comparison was carried out to show how the accuracy of the developed regression models depends on the presence of phenol red in the culture medium. The prediction of cancer cell numbers, particularly at the early stage of cancer, can be potentially applied as a reference feature in cytological analysis.

2. Materials and Methods

2.1. Cell Culture

Three cell lines including mouse skin fibroblast (L929), human prostate carcinoma (DU145), and human cervix adenocarcinoma (HeLa) acquired from American Type Culture Collection

(ATCC), Manassas, VA, USA, and used in this study. The procedure of the cells subculture and the spectroscopic system is similar to those described in our prior reports [25, 32]. The growth of the cells was performed in cell culture media which were Minimum Essential Medium 'MEM' (for L929 and DU145) and Dulbecco's Modified Eagle Medium 'DMEM' (for HeLa) with and without phenol red dye. The media were supplemented with fetal bovine serum 'FBS' (10%), sodium pyruvate (1 mM), and penicillin (100 units/mL)/streptomycin (100 µg/mL). The cell culture was preserved at 37 ℃ in an incubator with a 5% of CO_2 atmosphere. The cells viability test was carried out using a trypan blue dye exclusion technique by quantifying the number of viable cells using a hemocytometer. The cell lines were seeded in a 6-well cell culture plate with a total number of cells ranging from 50,000 to 275,000 with a step of 25,000 per well.

2.2. Spectra Acquisition

The absorbance spectra of each cell line cultured in a medium with and without phenol red were recorded using the Vis-NIR spectroscopy technique. The spectroscopy instrument used in the experiments is from Ocean Optics Inc. (Dunedin, Florida, USA). The spectroscopic system features and experimental setup are similar to those reported in our preceding studies [25, 32, 33]. The optical fiber was connected to a QE65000 spectrometer with a spectral sensitivity between 350.64 and 1131.24 nm. However, the wavelength region of 400-1100 nm was utilized in the entire analysis. Three wells of the 6-well cell culture were used, and triplicate spectra were acquired from each well to guarantee the consistency of the measurements and obtain the average spectrum for the entire analyses. The acquisition parameters for the spectrometer including integration time, scan to average, and boxcar width was 17 ms, 8, and 3, respectively. Spectra Suite Software (Ocean Optics) was used to acquire and analyze the recorded spectra.

2.3. Data Analysis

MLR was used to estimate the number of cells for L929, DU145, and HeLa cell lines. MLR analysis is a multivariate statistical technique utilized to model biological processes. The MLR explores the relationship between two or more independent variables and one dependent variable [34]. The MLR model ensures the selective extraction of specific information from highly overlapping NIR spectra. The use of an MLR model with fewer wavelengths is usually more robust since increasing the number of selected wavelengths may increase the specificity of the model to a particular dataset. Collinear data are inappropriate for MLR calibrations. The selection of the wavelengths depends on the coefficient of determination, standard error of the cross-validation method, and a significance level (α) [35]. Before performing the MLR, all X and Y variables were weighted to 1, and α of 0.05 was used. In addition, the selection of wavelengths is depending on the significance of each wavelength in the MLR model which is considered ≤ 0.05. Calibration and full cross-validation methods were used to develop and validate the performance of the MLR models, respectively. For the model validation, the leverage correction method was performed which is strictly similar to the full cross-validation method in the MLR [36]. Statistical characteristics that are used to show the performance of the constructed MLR models were the standard error of estimation (SEE), coefficient of determination (R^2), and a significance level (α) [37]. The SEE is defined as the root mean square error of estimation (RMSEE) corrected for bias; RMSEE is a measure of the distribution of the validation samples around the regression line. The value of $\alpha < 5\%$ in the MLR model indicates the model elucidates more of the deviations of the response variable compared to random phenomena. This means that the model is significant, and the smaller α, the higher significance of the model is. MLR modelling was performed on raw

spectral data without using any pre-processing using the Unscrambler Software (version 10.4, Camo Software AS, Oslo, Norway).

PLSR was used as another method in this study for quantification of the cells using the NIR region 680-1050 nm. To optimize the PLSR modelling, various spectral pre-processing techniques were examined such as standard normal variate (SNV) [38], multiplicative scatter correction (MSC) [39], smoothing using Savitzky-Golay [40] with 2nd order polynomial filter with 21- and 11-points window, 1st and 2nd Savitzky-Golay derivative transformation with 15 points, and orthogonal signal correction (OSC) [41]. It should be mentioned that before model development, the spectra were mean-centered. The validation was performed internally using the leave-one-out cross-validation method. As quality determination for respective models, the coefficient of determination (R^2) and standard error of cross-validation (SECV) were utilized. The optimal number of factors or latent variables (LVs) in the model was decided based on the local minimum of the SECV. The PLSR modelling was performed using a commercially available multivariate analysis software Pirouette v4.5 (Infometrix Inc., USA).

3. Results and Discussion

Figure 1 shows the example of the raw spectra (400-1050 nm) of one cell line which is DU145 cultured in MEM medium with and without phenol red dye. The spectra were labelled as DU145 PR1-10/DU145 XPR1-10 which means the DU145 was cultured in the medium with/without phenol red (PR/XPR) with the number of cells ranging from 50,000 to 275,000. The arrow in Figure 1 denotes the direction of the spectra with increasing the number of cells. Each spectrum is the average of nine measurements of spectra from three wells of the 6-well cell culture plate. The measurement of the spectrum from each well was repeated three times in different locations to ensure the repeatability of the measurement and eliminate the uncertainties due to

sampling. It is imperative to mention that the produced absorbance in Figure 1 resulted from both absorbance and scattering processes of the light and it depends on the physical and chemical properties of the samples [42]. Therefore, the observed absorbance can occur when light is scattered by a specific number of cells. As can be seen from Figure 1, there are two peaks in the visible and NIR regions. The predominant peak in the visible region (558 nm) denotes the basic form of phenol red [43, 44]. Given that the culture medium is composed of water, a distinctive absorbance peak is observed at NIR wavelength (≈ 965-985 nm) is referred to as the second overtone of the water absorbance peak [45]. It is evident that the absorbance of the samples in the NIR region is less affected or predisposed to the color of the medium.

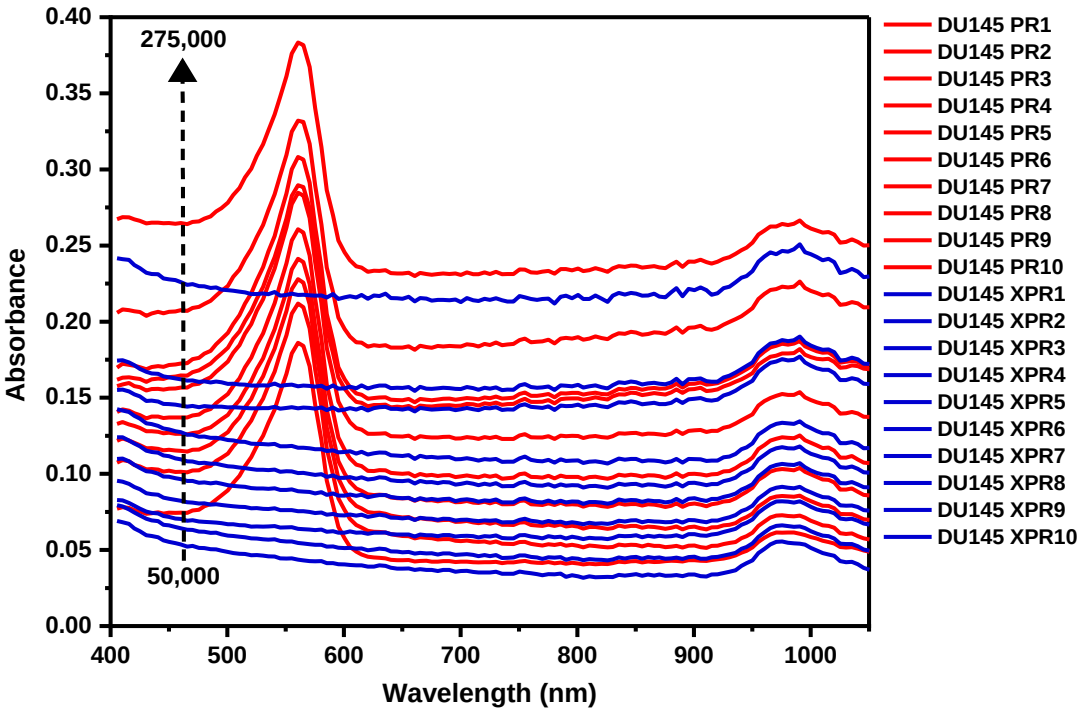


Figure 1. Vis-NIR absorbance spectra at region 400-1050 nm for DU145 cell line cultured in medium with phenol red (red color) and without phenol red (blue color) at the number of cells from 50,000 to 275,000.

NIR spectroscopy has the benefit of ease and expediency when applied to aqueous samples [46]. The use of NIR spectroscopy for cell cultivation has been recognized for the prediction of total or viable cell density and concentration of main metabolites. The peak absorbance values of the spectra at region 700-1050 nm for the three cell lines (L929, DU145, and HeLa) for 50,000 to 275,000 cells are presented in Supplementary Figure S1. In this figure, the bars represented the standard error of the mean (SE) of the peak intensity in the NIR region for each cell line at each number of cells. Each data point (mean $\pm$ SE) was calculated from nine recorded spectra that were obtained from three wells. Generally, it can be inferred that the peak absorbance values are consistent for the three cell lines. Nonetheless, there is a variation between the absorbance peak of the HeLa cell line grown in the presence and absence of phenol red in the culture medium at a high number of cells (250,000 and 275,000). This can be attributed to the relatively higher light scattering of HeLa cells compared to L929 and DU145 cells [25] and possibly due to a reduction in the emitted light at the detector. Light scattering minimizes the signal-to-noise ratio, thus intensifying the root mean square noise [47]. This may decrease the measurement precision of the absorbance spectra of the HeLa cell line. Furthermore, the variation in the peak absorbance in some number of cells may be because of errors by a person inexperienced with cell counting using the hemocytometer technique.

As reported in our previous work [33], the results showed that culturing different types of cells in medium with phenol red resulted in diminished interclass distance, and decreased the potential for discrimination between the three cell types (in PCA analysis). On the other hand, this has a benefit of easier quantification by creating general regression model that can be used for quantification of cells despite of their type. The aim of investigating the cells cultured in a medium with and without phenol red is to show the implication of the phenol red in the culture medium on

cells in the NIR region. The NIR absorbance spectra of the cells are less affected or predisposed to the color of the medium due to phenol red. Furthermore, it was showed that both water content and molecular structure of the cell culture (media and cells) are affected by staining.

In general, the pre-processing of the spectra is extensively used to eliminate/minimize the unwanted artefacts from the spectra. For PLSR/MLR models, various pre-processing methods and their combinations are frequently examined to find the optimum performance in terms of the minimum prediction error. In some cases, the scattering data might offer further merit associated with the physical composition of materials which could augment the behavior of the model [48]. As proved from the MLR and PLSR models in this study using NIR spectra at region 700-1050 nm of the three cell lines, the scattering may comprise vital details related to the feature of interest. Therefore, pre-processing techniques that diminish the scattering data (such as MSC and SNV) or their combination may result in badly processing models.

This study is aimed to build an algorithm model for quantification of the number of cells of L929, DU145, and HeLa cell lines. The MLR is appropriate for systems with a few variables [49] and is preferred on account of its non-complexity and easy operation [50]. The analysis was performed by applying MLR using the absorbance datasets of L929, DU145, and HeLa cell lines. The spectral data used was restricted to the 700-1050 nm wavelength range as the relative absorbance measurements outside this range were related to the color of the medium due to the presence of phenol red or too noisy. Before building the MLR models, the selected wavelengths (SWs) were obtained by the regression coefficients of PLSR models, and loadings plots of the PCA analysis. These wavelengths are considered to provide more contribution to the calibration of the MLR model [51].

In the case of including phenol red in the culture medium, the wavelengths that were selected for quantifying the number of the cells in the NIR region were 775, 841, 955 nm for L929 and for both DU145 and HeLa were 700, 801, and 976 nm. Individual (or local) MLR models were constructed using these wavelengths for the three cell lines. Besides, the local MLR models in the case of growing the cells in the media without phenol red were developed. The SWs were 736, 941, and 1015 nm for L929; 736, 775, 805, and 906 nm for DU145; 810, 861, 960, and 986 nm for the HeLa cell line. Table 1 summarizes the local MLR models and the statistics of the full cross validation (sometimes called Leave-One-Out "LOO" cross validation) [52] in the presence and absence of the phenol red. In addition, the predicted versus the actual number of cells using the calibration and cross-validation methods obtained from the local MLR models are shown in Figure 2(a)-(f) for L929, DU145, and HeLa cell lines. In this figure, the R^2 and standard errors of calibration and cross-validation (R_C^2, R_{CV}^2, SEC, and SEE) for each model are included.

Table 1. MLR local models and cross-validation statistics for each L929, DU145, and HeLa cell line to predict the number of cells in case of presence/absence of phenol red in culture media.

With phenol red											
L929				**DU145**				**HeLa**			
SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α
775, 841, 955	0.99	5,995	2.22E-08	700, 801, 976	0.99	7,498	7.01E-08	700, 801, 976	0.98	10,657	5.86E-07
Without phenol red											
L929				**DU145**				**HeLa**			
SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α
736, 941, 1015	1	4,771	1.29E-08	736, 775, 805, 906	1	3,363	2.33E-08	810, 861, 960, 986	0.95	17,442	9.44E-05

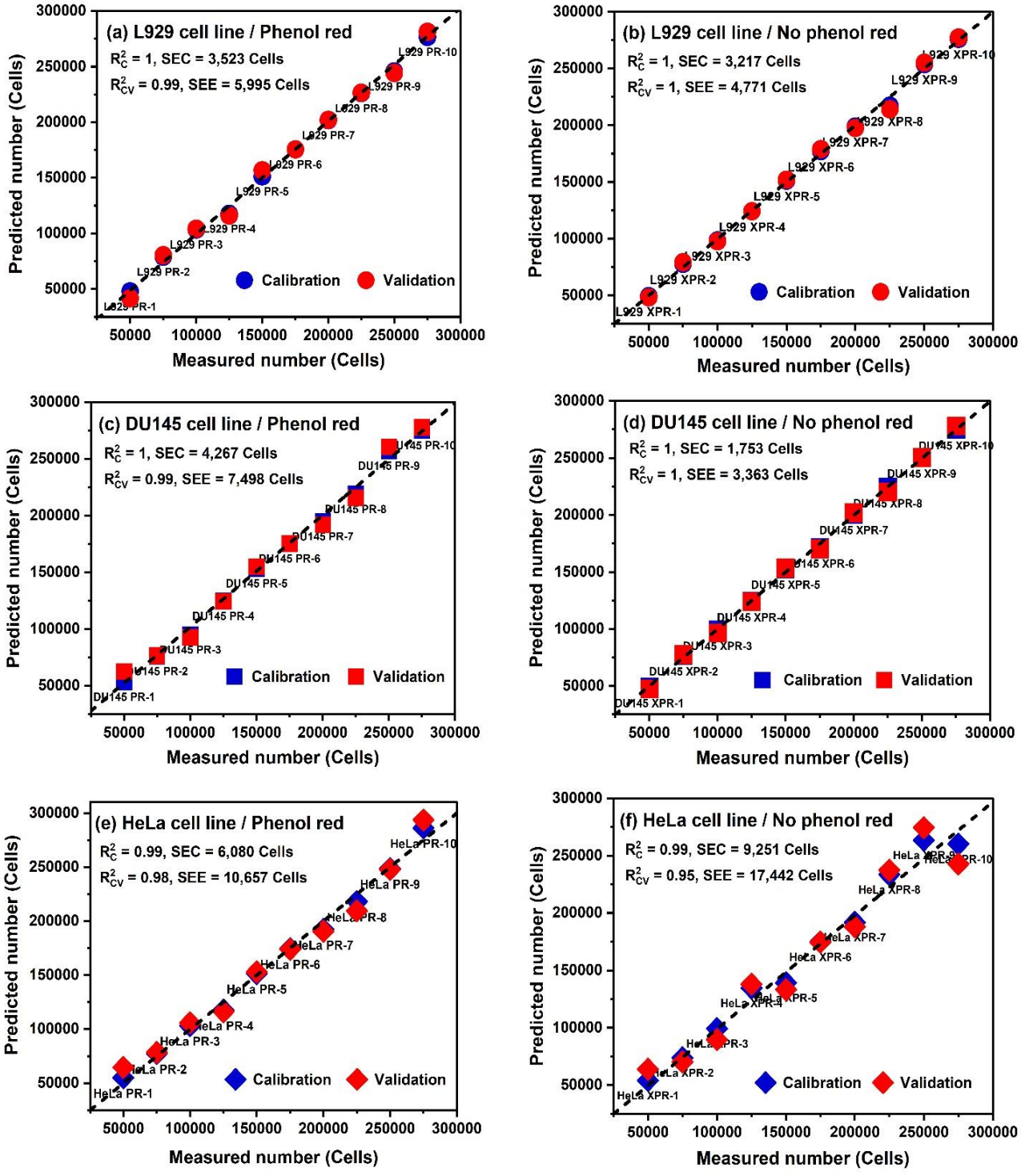


Figure 2. Predicted versus measured values of calibration (red) and cross-validation (blue) results of local MLR models for quantification the cells: (a) and (b) for L929, (c) and (d) for DU145, and (e) and (f) for HeLa cell line in case of presence/absence phenol red in culture media, respectively.

A general (or global) MLR model for estimating the number of cells of the three cell lines together (L929, DU145, and HeLa) in the presence (using SWs: 736, 775, and 805 nm) and absence (using SWs: 745, 810, 861, 900, 960, 986, and 1010 nm) of the phenol red in the culture media was constructed. The samples were designated as "PR" and "XPR" which means the presence and absence of phenol red, respectively. Table 2 shows the global MLR models and the

Table 2. MLR global models and cross-validation statistics for L929, DU145, and HeLa cell lines to predict the number of cells in case of presence/absence of phenol red in culture media.

With phenol red			
SWs (nm)	R^2_{CV}	SEE (Cells)	α
736, 775, 805	0.96	14,661	3.47E-20
Without phenol red			
SWs (nm)	R^2_{CV}	SEE (Cells)	α
745, 810, 861, 900, 960, 986, 1010	0.93	19,138	1.22E-14

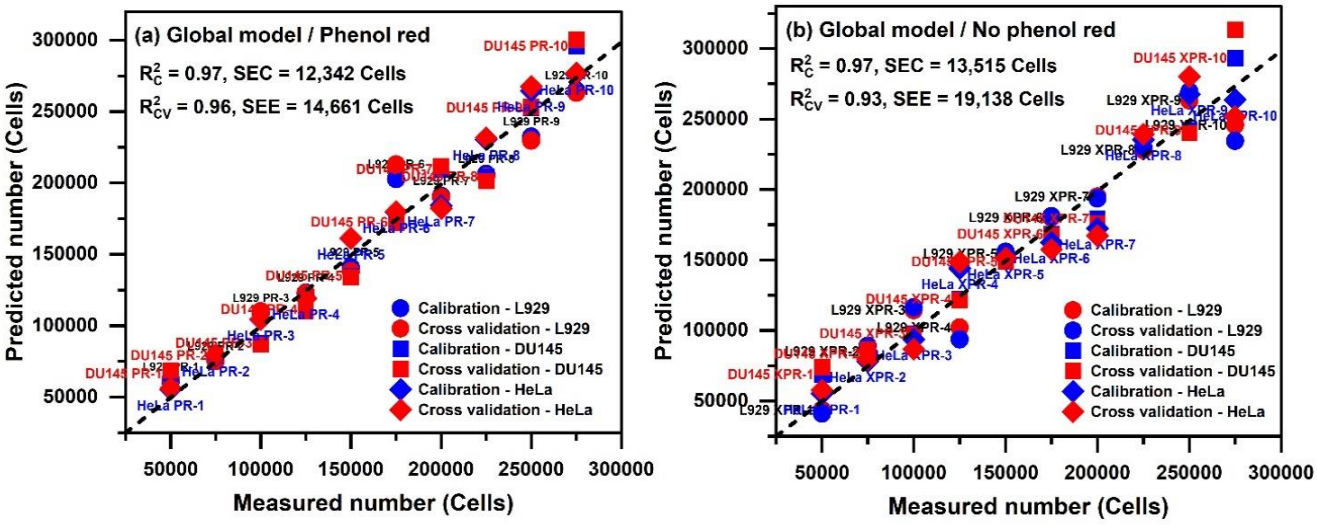


Figure 3. Predicted versus measured values of calibration (red) and cross-validation (blue) results of global MLR models for quantification of the L929, DU145, and HeLa cells grown in the medium: (a) with phenol red and (b) without phenol red.

statistics of the cross-validation method in the presence and absence of the phenol red. Figure 3 displays the global MLR models for predicting the number of cells. The obtained values of R^2_{CV} and SEE of the MLR models suggested its good predictive capacity and the appropriateness of the selected wavelengths in the spectral region 700-1050 nm without any spectral pre-treatments.

The PLSR modelling was performed in two ways: firstly, a local PLSR model was created for each cell line separately (L929, DU145, and HeLa) that is best adapted for a specific cell line. Secondly, a global PLSR model was developed for quantification of the number of cells irrespective of the cell line type. In both cases, models were built separately for the cells grown in media with and without phenol red to discern the effects of phenol red on cell culture and quantification. Quantitative modelling was first performed using PLSR analysis on datasets of spectra for each cell line separately in order to build local models assuming that the accuracy of prediction would be the highest and allow assessment of effects of phenol red treatment. Table 3 presented the PLSR model and validation statistics of testing different pre-processing techniques for local PLSR models in both cases with and without the presence of phenol red treatments. From Table 3, it can be observed that the removal of additive and multiplicative effects (scatter correction techniques: SNV, MSC, and $1^{st}/2^{nd}$ derivatives) led to poorer model performance. This means that the baseline variations contain significant information for the quantification of cells, which is in agreement with previous research reports about the quantifications of fat (globules), protein (micelles), and somatic cells [53, 54]. Based on the quality measures for respective PLSR models, it can be established that the best accuracy was obtained with the OSC pre-processing method using one component, irrespective of phenol red treatment and cell line type as shown in Table 3. The validation results showed R^2 higher than 0.99 using the OSC pre-processed spectra. The treatment of phenol red improved the accuracy of cells number prediction in the case of the

cancer cells (DU145: SECV = 250 cells and HeLa: SECV = 271 cells). However, for L929 normal cells, the accuracy of the PLSR models was enhanced in the case of not including the phenol red in the culture medium (SECV = 24 cells). In addition, the accuracy of quantification was poorest for HeLa cells, and the best in the case of L929 cells. Figure 4 shows the results of calibration and leave-one-out cross-validation methods of the local PLSR models developed using the pre-processing by OSC (one component) for mean-centered spectra. This figure shows a good agreement between the measured and predicted number of cells by PLSR model. Figure 5

Table 3. Local PLSR models and cross-validation statistics for each L929, DU145, and HeLa cell line to predict the number of cells in case of presence/absence phenol red in culture media.

Case	**Pre-processing method**	**L929**			**DU145**			**HeLa**		
		LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)
With phenol red	None	5	0.99	12,059	3	0.99	11,685	4	0.99	8,064
	Smoothing (11 points)	2	0.99	12,517	4	0.99	9,348	3	0.99	11,275
	SNV	4	0.89	34,022	6	0.86	37,709	5	0.79	44,570
	MSC	5	0.58	73,835	2	0.83	41,045	5	0.8	43,929
	1st Derivative (15 points)	7	0.69	55,121	6	0.93	27,399	6	0.89	34,213
	2nd Derivative (15 points)	3	0.76	46,852	3	0.96	21,256	2	0.78	45,231
	OSC	**2**	**1**	**137**	**1**	**1**	**250**	**1**	**1**	**271**
Case	**Pre-processing method**	**L929**			**DU145**			**HeLa**		
		LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)
Without phenol red	None	2	0.99	5,570	5	0.96	23,077	2	0.96	21,032
	Smoothing (21 points)	2	0.99	5,625	4	0.97	19,351	2	0.96	20,993
	SNV	2	0.78	45,562	3	0.96	18,939	1	0.78	46,253
	MSC	2	0.78	45,525	3	0.96	19,008	1	0.76	47,500
	1st Derivative (15 points)	2	0.62	60,557	3	0.97	18,751	1	0.78	44,847
	2nd Derivative (15 points)	1	0.68	54,716	1	0.92	27,577	3	0.74	48,440
	OSC	**3**	**1**	**24**	**1**	**1**	**1,364**	**1**	**0.99**	**7,444**

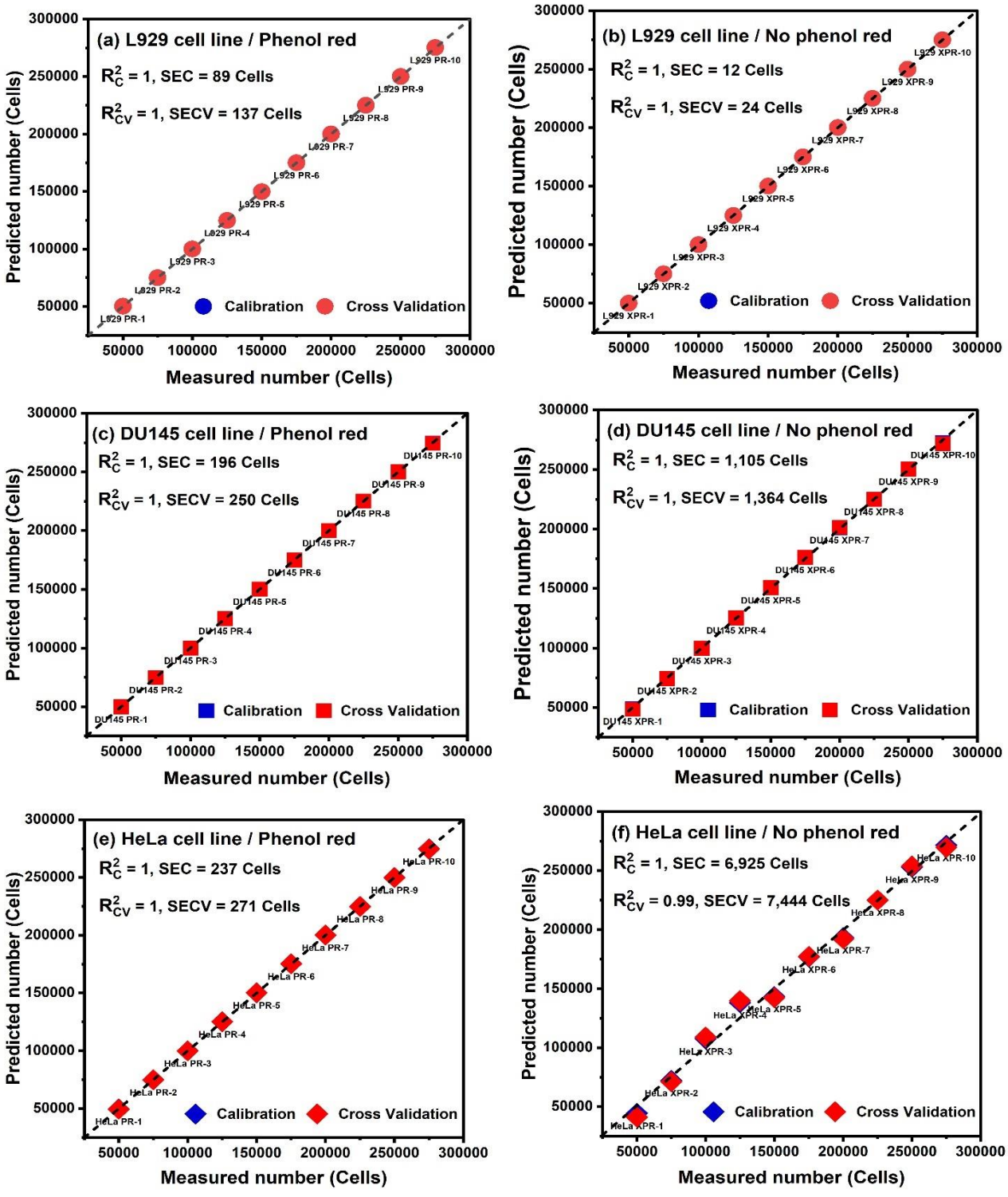


Figure 4. Predicted versus measured values of calibration (red) and cross-validation (blue) results of local PLSR models for quantification of cells: (a) and (b) for L929, (c) and (d) for DU145, and (e) and (f) for HeLa cell line in case of presence/absence phenol red in culture media, respectively.

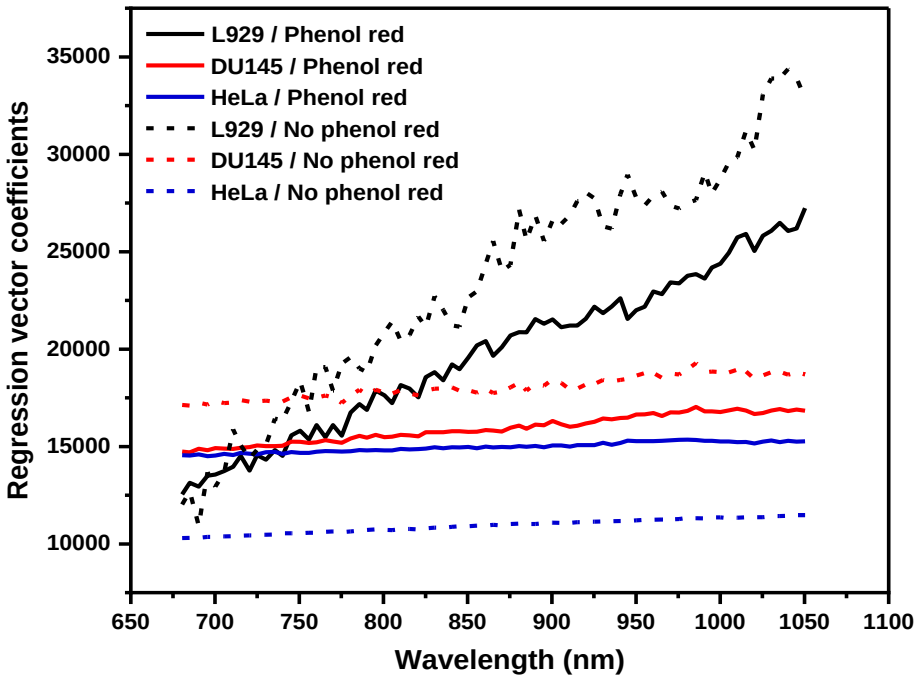


Figure 5. Regression vector coefficients of local PLSR models obtained from the best accuracy of prediction of the number of cells of L929, DU145, and HeLa cell lines in case of presence/absence of phenol red in culture media.

shows the regression vector coefficients of the PLSR models developed from Figure 4 for each cell line. The lack of any marked peaks validates the fact that the scatter plays the important role in the PLSR modelling instead of the absorbance [52]. It can be observed from Figure 5 that the cells' quantification is related to the variations of the baseline, in other words, the quantification of the cells is based on the effects of light scattering by the cells. In addition, from Figure 5, in the case of both DU145 and HeLa cell lines (cancer cells), the baseline effects are mostly additive resulting in the shift of the baseline of the spectra. However, in the case of the L929 cell line (normal cells), there are also multiplicative effects resulting in the change of the slope of the baseline. Interestingly, the use of phenol red influences the scattering of light by cells which is attested by the shifting in the regression vectors, but the shape is quite preserved. It can be deduced that the building of the local PLSR models for each cell line resulted in high accuracy of the

quantification of the cells. The quantification of the cells was better when L929 cells were grown in the medium without phenol red, while on the opposite, the quantification accuracy for DU145 and HeLa extremely improved when the phenol red was included in the culture media. These findings demonstrate the influence of phenol red on light-scattering properties, which are different and dependent on the cell type.

Two global PLSR models in the case of the presence and absence of phenol red were developed that allowed quantification of cells irrespective of the cell type. Various pre-processing techniques were also used in order to extract the information related to the number of cells and enhance the performance of the global model. The results of the cross-validation method of the developed global PLSR models using different pre-processing methods are summarized in

Table 4. Global PLSR models and cross-validation statistics for L929, DU145, and HeLa cell lines to predict the number of cells in case of presence/absence of phenol red in culture media.

Case	Pre-processing method	LVs	R^2_{CV}	SECV (Cells)
With phenol red	None	5	0.98	13,007
	Smoothing	7	0.98	13,694
	SNV	3	0.73	50,343
	MSC	3	0.62	60,740
	1st Derivative	5	0.88	34,176
	2nd Derivative	7	0.75	49,802
	OSC	**1**	**0.99**	**7,754**
Case	**Pre-processing method**	**LVs**	R^2_{CV}	**SECV (Cells)**
Without phenol red	None	5	0.93	26,577
	Smoothing	5	0.94	25,447
	SNV	3	0.79	43,826
	MSC	3	0.8	43,606
	1st Derivative	5	0.88	34,380
	2nd Derivative	2	0.82	41,279
	OSC	**1**	**0.97**	**17,008**

Table 4. It is clear from Table 4 that also for both global models, OSC pre-processing using one component produced the best accuracy of quantification and resulted in a simple model with only one latent variable. Compared to the established local models for each cell line, the global models showed a significant decrease in the accuracy of predicting the number of cells. This implies that local modelling considers the specific characteristics of each cell line which is a better choice in quantitative modelling. Furthermore, from Table 4, the phenol red in the cultured medium resulted in improved accuracy of cells' quantification. The relationship between the measured and predicted number of cells using the global PLSR models developed using the pre-processed spectra by the OSC method for both cases with and without phenol red treatments is shown in Figures 6(a) and (b), respectively. It can be inferred that the phenol red has effects on the cells, in the terms of diminishing the differences between the individual cell lines. This is in good agreement with the findings of our preceding study, wherein in the case of discrimination between the three cell lines, cells cultured in media including phenol red showed diminished class distances in discriminative models [33]. As can be seen from Figure 6(a), it is evident that phenol red in the cell culture media led to a diminishment of the error and decreased differences between the three cell lines. This confirms the influence of the phenol red on the characteristics of the cells and their subsequent interaction with near-infrared light. On the other hand, for the cells grown in the media without phenol red as presented in Figure 6(b), there is a small error between the measured and predicted number of cells at a small number of cells. Whereas, at a higher number of cells, the error increases and the distinction between the cell lines is more pronounced. The regression vector coefficients of the PLSR global models of the three cell lines (L929, DU145, and HeLa) are presented in Figure 7. Apparently, the regression vectors of the global PLSR model established for the case of cells grown in media with phenol red (solid line) show a higher magnitude of coefficients. This can be

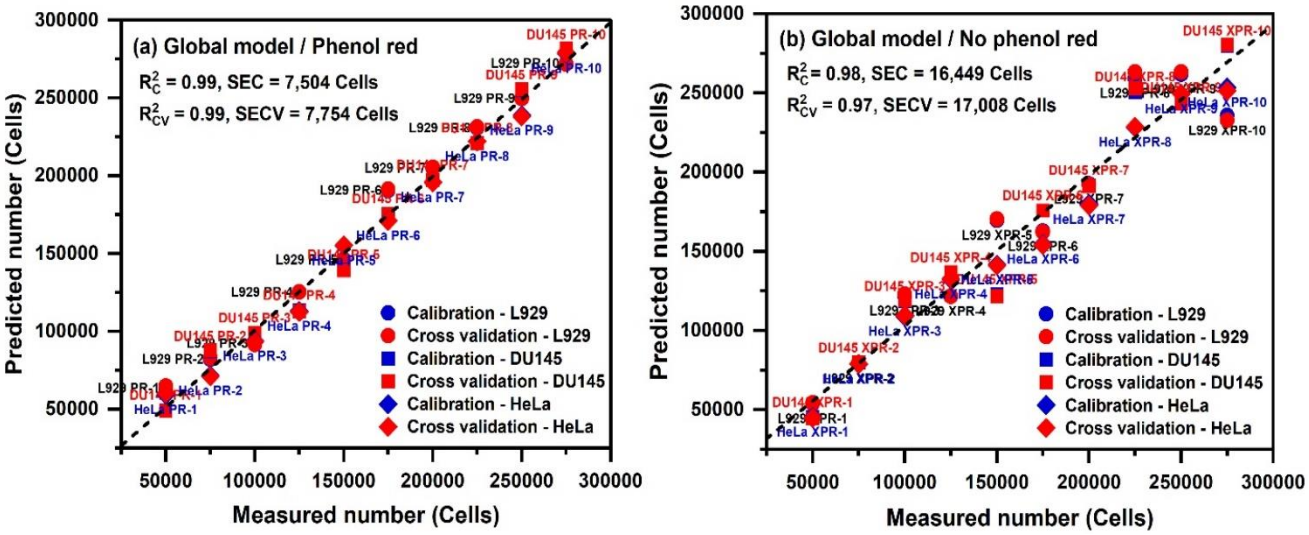


Figure 6. Predicted versus measured values of calibration (red) and cross-validation (blue) results of global PLSR models for quantification of the L929, DU145, and HeLa cells grown in the medium: (a) with phenol red and (b) without phenol red.

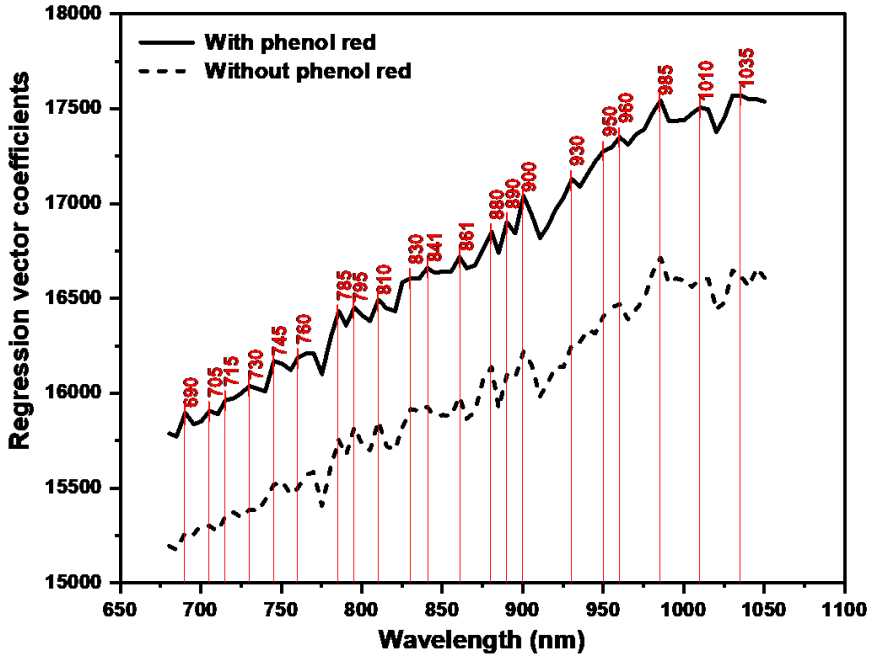


Figure 7. Regression vector coefficients s of global PLSR models for quantification of L929, DU145, and HeLa cells grown in the media with and without phenol red.

attributed to the increased scattering of light by the cells. However, the shape of the regression coefficients is very similar for both models. This revealed that both global models are based on

variations of the spectral baseline in connection to the number of cells. Despite the dominant effect of scattering, absorbance bands can be observed in the regression coefficients and almost comparable absorbance bands characterize the regression coefficients of the two global models.

In the NIR spectral region, particularly at the region below 1000 nm, the scattering strongly dominates over absorption. It was reported in various works that this region can be used as a source for quantitative information on light-scattering of colloidal particles [52, 55-57] which can explain the findings acquired in this work. The main factors affecting the scattering of a single particle or a collection of particles are the refractive index difference between the particles and the matrix, the size of the particle, and the wavelength of the incident light [58]. It is most likely that the refractive index of the matrix (the medium in which the cells were grown, and the highest component of the medium is water) is affected by the presence of phenol red.

For more investigation, we examined the difference in influential variables between the created local PLSR models depending on the presence (PR) and absence (XPR) of phenol red in the media where the cells were grown. For this purpose, the regression vector coefficients of the corresponding models were detrended to eliminate the slope, and then the difference between the regression vector coefficients of PR and XPR models was calculated. The plots of regression vector coefficients after detrending and the result of their subtraction are presented in Figures 8-10 for L929, DU145, and HeLa PLSR models, respectively. Interestingly, the highest differences in regression coefficients in all models are located in the same region which is known as the 2nd overtone of water (approximately around 900-1000 nm). The largest differences in this region can be observed at bands: 915 and 945 nm for L929 cells, 955 and 986 nm for DU145 cells, and 930, 945, 970, and 976 nm for HeLa cells. Obviously, there is a difference in these particular bands in the terms of their importance as influential variables for building the PLSR model. This difference

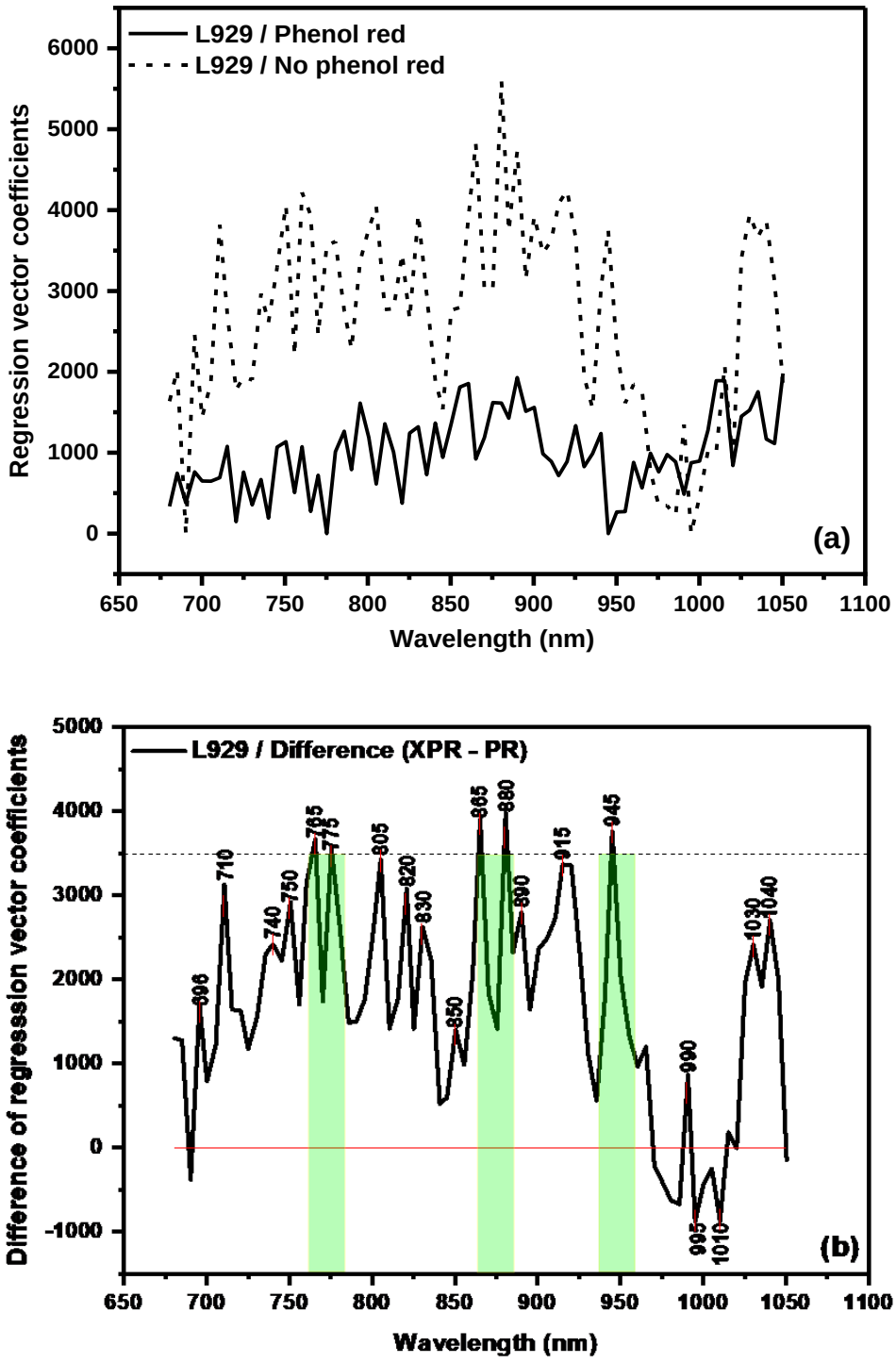


Figure 8. Detrended regression vector coefficients of PLSR models for L929 cells quantification built using spectral data from cells in media with phenol red (PR) and without phenol red (XPR) (a) and their difference (b).

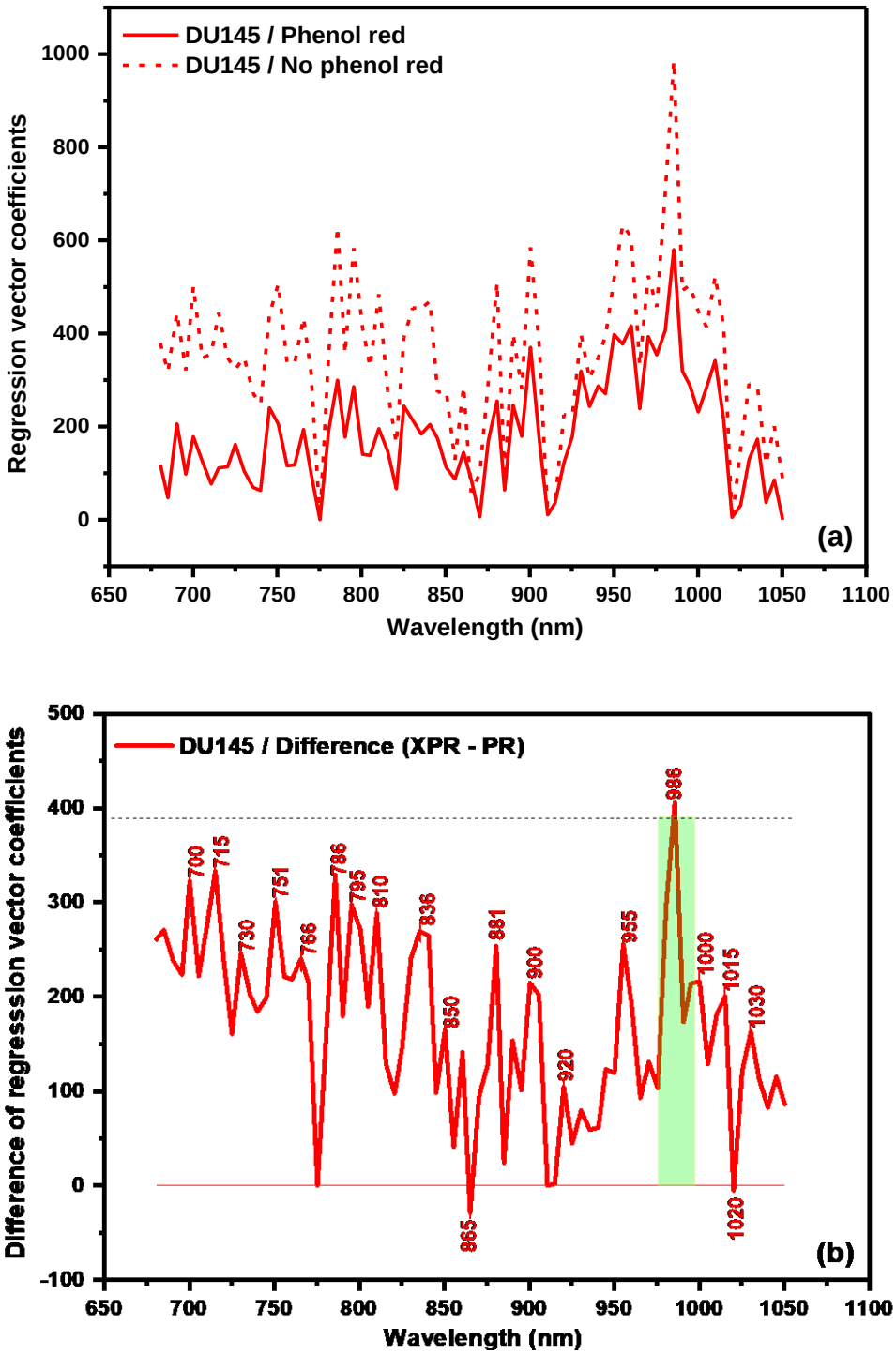


Figure 9. Detrended regression vector coefficients of PLSR models for DU145 cells quantification built using spectral data from cells in media with phenol red (PR) and without phenol red (XPR) (a) and their difference (b).

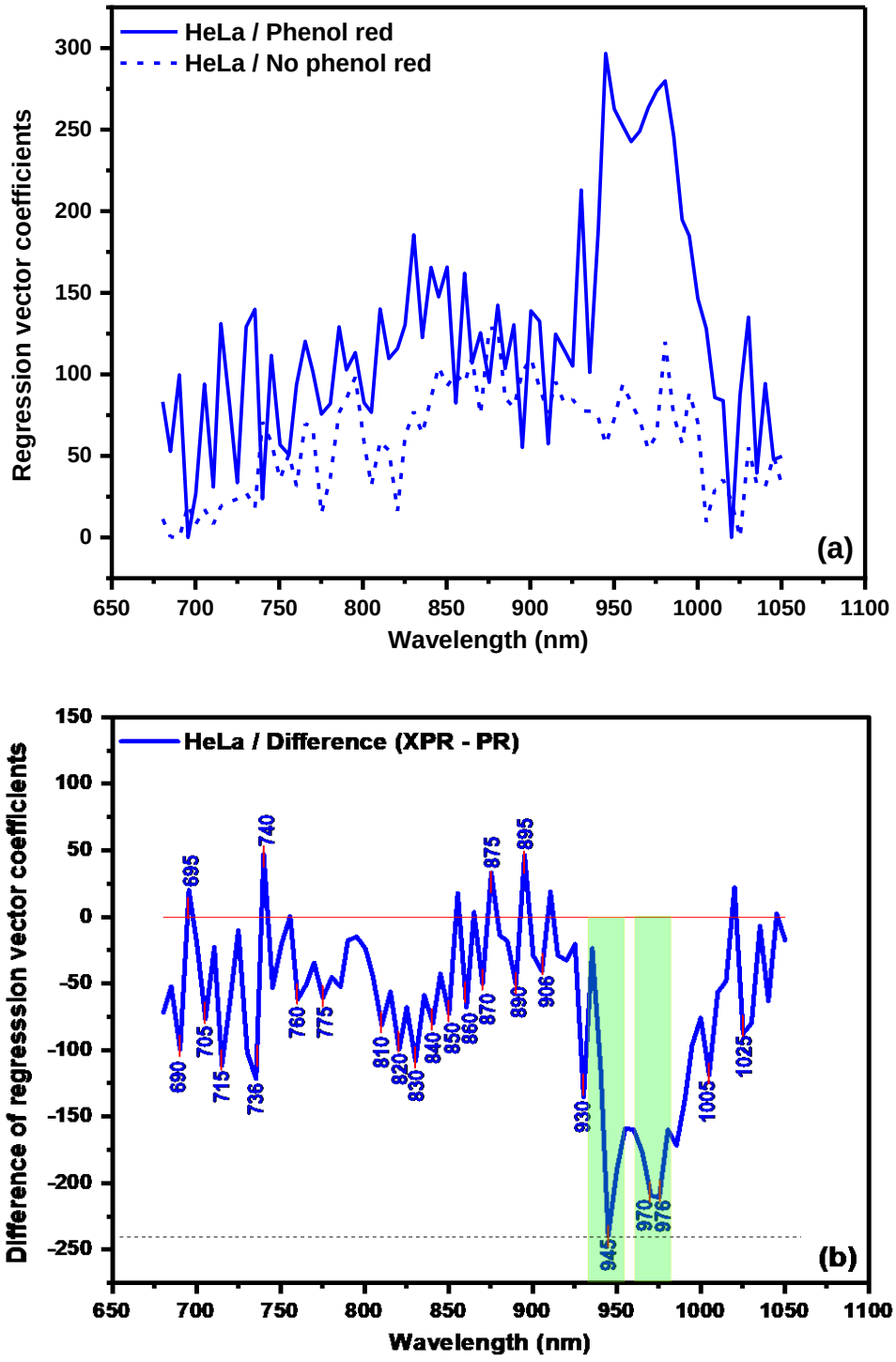


Figure 10. Detrended regression vector coefficients of PLSR models for HeLa cells quantification built using spectral data from cells in media with phenol red (PR) and without phenol red (XPR) (a) and their difference (b).

depends on whether the cells were grown in the absence or presence of phenol red and is specific for each cell type. The assignments of the bands that are of importance (the highest bands) for the accuracy of the PLSR models depending on the presence of phenol red in the culture media are given in Table 5. Since the bands' positions were in the 2nd overtone region of water, their assignments were found based on the calculated position in the 1st overtone of water as presented in Table 5. The bands shown in Table 5 are the 2nd overtones of the well-known Water Matrix Coordinates "WAMACs" in aquaphotomics (water absorbance bands attributable to vibrations of the particular water molecular species) [59]. This confirms that the presence of phenol red has an influence on the cells' culture media by changing the water structure. Further, considering that different bands are affected by the type of the cells, the water structure of the media is a result of both influences of cell metabolites and phenol red. As a result, for the DU145 cells, the specific bands 955 and 986 nm, i.e., water dimers and tetramers are affected; in the case of HeLa cells are water trimers, solvation shells, free and quasi-free water molecules, while in the case of L929

Table 5. Assignments of the bands found to be of importance for cells quantification accuracy using PLSR analysis, depending on the presence of phenol red in the cell media.

Band position in 2nd overtone of water (nm)	**Calculated position in 1st overtone of water (1300 - 1600 nm)**	**Assignment / WAMACs [59-61]**
915	1372.5	C3 – $\nu_1 + \nu_3$: Water symmetrical stretching vibration and water asymmetric stretching vibration
930	1395	C5 – Water molecules confined in the local field of ions
945	1417.5	C5 – Free water molecules, S_0
955	1432.5	C6 – S1: Water molecules with 1 hydrogen bond
970	1455	C8 – Water solvation shell, $OH^-(H_2O)_{4,5}$
976	1464	C9 – Water molecules with 2 hydrogen bonds, S_2
986	1479	C10 – Water molecules with 3 hydrogen bonds, S_3

cells, in particular free water molecules. It is important to point out that in the case of L929 cells, there is a small difference between the regression vector coefficients at hydrogen-bonded water wavelengths ($>$ 945 nm), and this makes a stark difference when compared to DU145 and HeLa cells where the largest differences were observed exactly in hydrogen-bonded water. Since in all cases the phenol red caused changes in the molecular structure of water, the refractive index must also change. This can be explained by the refractive index dependence on the molecular structure [62] of the matrix, resulting in different scattering properties of each cell line media and different quantification outcomes.

The results of this study should also bring more realization to the so-called "Observer Effect", an important and frequent experience in biomedical research, where the tools used to facilitate observation may exert unrecognized and often undesired effects on the underlying biology [63]. While the phenol red is routinely used in laboratory practice, its effects on the system the researchers are working with are still unrecognized. This study showed the effective treatment of the phenol red dye towards better quantification of the cultured cells. Furthermore, the molecular structure of water is not yet recognized as an important experimental variable and parameter that should be considered in research studies. It was shown here that the molecular structure of water was altered by the presence of phenol red in the cells' culture medium. The findings of this study, therefore, enhance awareness of these issues and revealed the informative power of NIR spectroscopy and aquaphotomics in this regard.

4. Conclusion

This work presented the application of NIR absorbance spectroscopy data to quantify the number of cultured cells. Three cell lines including cancerous cells (DU145 and HeLa) and normal cells (L929) were used. The number of cells for each cell line was ranging from 50,000 to 275,000

with an interval of 25,000. The Vis-NIR absorbance spectra (400-1100 nm) were measured for each number of cells cultured in media with and without phenol red. To quantify the number of cells for the three cell lines, the local and global regression models using multiple linear regression (MLR) and partial least squares regression (PLSR) were developed in the NIR region (680-1050 nm). It was revealed that the MLR and PLSR models for the quantification of the cells produced high prediction accuracy with $R^2 \geq 93\%$. It was noticed that the accuracy of the local and global PLSR models, especially for cancer cells was higher in the case of cultured cells in media including phenol red dye. This can be an important finding in this work which showed the implication of phenol red in cell culture media towards the quantification of the number of cancerous cells. The proposed method in this study can be considered a novel and rapid approach for quantifying the number of cancerous cells.

Acknowledgment

The authors with gratitude acknowledge the financial support of this research by the Ministry of Higher Education Malaysia–Fundamental Research Grant Scheme (Grant No. FRGS/1/2020/STG07/USM/02/8).

References

1. Sorvina, A., et al., *Lipid profiles of prostate cancer cells.* Oncotarget, 2018. **9**(85): p. 35541.

2. Loud, J.T. and J. Murphy. *Cancer screening and early detection in the 21st century*. in *Seminars in oncology nursing*. 2017. Elsevier.

3. Vembadi, A., A. Menachery, and M.A. Qasaimeh, *Cell cytometry: Review and perspective on biotechnological advances.* Frontiers in bioengineering and biotechnology, 2019. **7**: p. 147.

4. Friberg, S. and S. Mattson, *On the growth rates of human malignant tumors: implications for medical decision making.* Journal of surgical oncology, 1997. **65**(4): p. 284-297.

5. Frangioni, J.V., *New technologies for human cancer imaging.* Journal of clinical oncology, 2008. **26**(24): p. 4012.

6. Shern Khoo, A.B., et al., *Comparative Analyses of Tumour Volume Doubling Times for Periocular and Non-periocular Head and Neck Basal Cell Carcinomas.* Acta dermato-venereologica, 2019. **99**(13): p. 1266-1269.

7. O'Brien, J., H. Hayder, and C. Peng, *Automated quantification and analysis of cell counting procedures using ImageJ plugins.* Journal of visualized experiments: JoVE, 2016(117).

8. Ongena, K., et al., *Determining cell number during cell culture using the Scepter cell counter.* Journal of visualized experiments: JoVE, 2010(45).

9. Brown, J.Q., et al., *Advances in quantitative UV–visible spectroscopy for clinical and pre-clinical application in cancer.* Current opinion in biotechnology, 2009. **20**(1): p. 119-131.

10. Solano, R.P., et al., *An experimental and theoretical approach to the study of the photoacoustic signal produced by cancer cells.* AIP Advances, 2012. **2**(1): p. 011102.

11. Krediet, C.J., et al., *Rapid, precise, and accurate counts of symbiodinium cells using the guava flow cytometer, and a comparison to other methods.* PLoS One, 2015. **10**(8): p. e0135725.

12. Blasi, T., et al., *Label-free cell cycle analysis for high-throughput imaging flow cytometry.* Nature communications, 2016. **7**(1): p. 1-9.

13. Grishagin, I.V., *Automatic cell counting with ImageJ.* Analytical biochemistry, 2015. **473**: p. 63-65.

14. Mölder, A., et al., *Non-invasive, label-free cell counting and quantitative analysis of adherent cells using digital holography.* Journal of microscopy, 2008. **232**(2): p. 240-247.

15. Talebian, S. and M. Javanmard, *Compact and automated particle counting platform using smartphone-microscopy.* Talanta, 2021. **228**: p. 122244.

16. Sun, T. and H. Morgan, *Single-cell microfluidic impedance cytometry: a review.* Microfluidics and Nanofluidics, 2010. **8**(4): p. 423-443.

17. von Bartheld, C.S., J. Bahney, and S. Herculano-Houzel, *The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting.* Journal of Comparative Neurology, 2016. **524**(18): p. 3865-3895.

18. Trunfio, N., et al., *Characterization of mammalian cell culture raw materials by combining spectroscopy and chemometrics.* Biotechnology progress, 2017. **33**(4): p. 1127-1138.

19. Oshima, Y., et al., *Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy.* Journal of biomedical optics, 2010. **15**(1): p. 017009.

20. Dai, W.Y., S. Lee, and Y.C. Hsu, *Discrimination between oral cancer and healthy cells based on the adenine signature detected by using Raman spectroscopy.* Journal of Raman Spectroscopy, 2018. **49**(2): p. 336-342.

21. Shin, H., et al., *Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes.* ACS nano, 2020. **14**(5): p. 5435-5444.

22. Chan, J.W., et al., *Nondestructive identification of individual leukemia cells by laser trapping Raman spectroscopy.* Analytical chemistry, 2008. **80**(6): p. 2180-2187.

23. Talari, A., et al., *Raman spectroscopic analysis differentiates between breast cancer cell lines.* Journal of Raman Spectroscopy, 2015. **46**(5): p. 421-427.

24. Del Mistro, G., et al., *Surface-enhanced Raman spectroscopy of urine for prostate cancer detection: a preliminary study.* Analytical and Bioanalytical Chemistry, 2015. **407**(12): p. 3271-3275.

25. Abd Ghani, K., et al., *VIS–NIR spectral signature and quantitative analysis of HeLa and DU145 cell line.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019. **222**: p. 117241.

26. Sakudo, A., *Near-infrared spectroscopy for medical applications: Current status and future perspectives.* Clinica Chimica Acta, 2016. **455**: p. 181-188.

27. Geladi, P. and B.R. Kowalski, *Partial least-squares regression: a tutorial.* Analytica chimica acta, 1986. **185**: p. 1-17.

28. Martens, H. and M. Martens, *Multivariate analysis of quality: an introduction*. 2001: John Wiley & Sons.

29. Geladi, P. and E. Dåbakk, *Computational methods and chemometrics in near infrared spectroscopy.* 2017.

30. Shahmirani, S., E. Vasheghani Farahani, and J. Ghasemi, *Development of a model to predict partition coefficient of organic pollutants in cloud point extraction process.* Annali di Chimica: Journal of Analytical, Environmental and Cultural Heritage Chemistry, 2006. **96**(5-6): p. 327-337.

31. Carrascal, L.M., I. Galván, and O. Gordo, *Partial least squares regression as an alternative to current regression methods used in ecology.* Oikos, 2009. **118**(5): p. 681-690.

32. Amran, E.N., et al., *Potential colorimetric detection of cancer cells using Phenol Red.* Photodiagnosis and photodynamic therapy, 2019. **27**: p. 380-384.

33. Raypah, M.E., et al., *Integration of near-infrared spectroscopy and aquaphotomics for discrimination of cultured cancerous cells using phenol red.* Chemometrics and Intelligent Laboratory Systems, 2022. **227**: p. 104611.

34. Jobson, J., *Multiple linear regression*, in *Applied multivariate data analysis*. 1991, Springer. p. 219-398.

35. Berntsson, O., G. Zackrisson, and G. Östling, *Determination of moisture in hard gelatin capsules using near-infrared spectroscopy: applications to at-line process control of pharmaceutics.* Journal of pharmaceutical and biomedical analysis, 1997. **15**(7): p. 895-900.

36. Esbensen, K.H., et al., *Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design*. 2002: Multivariate Data Analysis.

37. Farhadi, S., et al., *Modeling of paclitaxel biosynthesis elicitation in Corylus avellana cell culture using adaptive neuro-fuzzy inference system-genetic algorithm (ANFIS-GA) and multiple regression methods.* PloS one, 2020. **15**(8): p. e0237478.

38. Barnes, R., M.S. Dhanoa, and S.J. Lister, *Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.* Applied spectroscopy, 1989. **43**(5): p. 772-777.

39. Geladi, P., D. MacDougall, and H. Martens, *Linearization and scatter-correction for near-infrared reflectance spectra of meat.* Applied spectroscopy, 1985. **39**(3): p. 491-500.

40. Savitzky, A. and M.J. Golay, *Smoothing and differentiation of data by simplified least squares procedures.* Analytical chemistry, 1964. **36**(8): p. 1627-1639.

41. Wold, S., et al., *Orthogonal signal correction of near-infrared spectra.* Chemometrics and Intelligent laboratory systems, 1998. **44**(1-2): p. 175-185.

42. Blanco, M., et al., *Effect of data preprocessing methods in near-infrared diffuse reflectance spectroscopy for the determination of the active compound in a pharmaceutical preparation.* Applied Spectroscopy, 1997. **51**(2): p. 240-246.

43. Guo, Y., et al., *A pH-responsive colorimetric strategy for DNA detection by acetylcholinesterase catalyzed hydrolysis and cascade amplification.* Biosensors and Bioelectronics, 2017. **94**: p. 651-656.

44. Yang, Y., et al., *A pH-responsive bioassay for paper-based diagnosis of exosomes via mussel-inspired surface chemistry.* Talanta, 2019. **192**: p. 325-330.

45. Ali, J., et al., *Near infrared spectroscopy and imaging to probe differences in water content in normal and cancer human prostate tissues.* Technology in cancer research & treatment, 2004. **3**(5): p. 491-497.

46. Liu, K.-Z., et al., *Quantitative determination of serum LDL cholesterol by near-infrared spectroscopy.* Vibrational spectroscopy, 2005. **38**(1-2): p. 203-208.

47. Petiot, E., et al., *In situ quantification of microcarrier animal cell cultures using near-infrared spectroscopy.* Process Biochemistry, 2010. **45**(8): p. 1427-1431.

48. Mishra, P., et al., *Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction.* Talanta, 2021. **229**: p. 122303.

49. Martens, H. and T. Naes, *Multivariate calibration*. 1992: John Wiley & Sons.

50. Escandar, G.M., et al., *A review of multivariate calibration methods applied to biomedical analysis.* Microchemical Journal, 2006. **82**(1): p. 29-42.

51. Wang, A. and L. Xie, *Technology using near infrared spectroscopic and multivariate analysis to determine the soluble solids content of citrus fruit.* Journal of Food Engineering, 2014. **143**: p. 17-24.

52. Bogomolov, A., et al., *Quantitative determination of fat and total protein in milk based on visible light scatter.* Food Chemistry, 2012. **134**(1): p. 412-418.

53. Tsenkova, R., et al., *Somatic cell count determination in cow's milk by near-infrared spectroscopy: A new diagnostic tool.* Journal of animal science, 2001. **79**(10): p. 2550-2557.

54. Bogomolov, A. and A. Melenteva, *Scatter-based quantitative spectroscopic analysis of milk fat and total protein in the region 400–1100 nm in the presence of fat globule size variability.* Chemometrics and Intelligent Laboratory Systems, 2013. **126**: p. 129-139.

55. Bogomolov, A., et al., *Reference-free spectroscopic determination of fat and protein in milk in the visible and near infrared region below 1000 nm using spatially resolved diffuse reflectance fiber probe.* Talanta, 2017. **167**: p. 563-572.

56. Bogomolov, A., A. Melenteva, and D.J. Dahm, *Fat globule size effect on visible and shortwave near infrared spectra of milk.* Journal of Near Infrared Spectroscopy, 2013. **21**(5): p. 435-440.

57. Surkova, A., et al., *Towards an optical multisensor system for dairy: Global calibration for fat analysis in homogenized milk.* Microchemical Journal, 2019. **149**: p. 104012.

58. Dahm, D.J., *Explaining some light scattering properties of milk using representative layer theory.* Journal of Near Infrared Spectroscopy, 2013. **21**(5): p. 323-339.

59. Tsenkova, R., *Aquaphotomics: dynamic spectroscopy of aqueous and biological systems describes peculiarities of water.* Journal of Near Infrared Spectroscopy, 2009. **17**(6): p. 303-313.

60. Kojić, D., et al., *Water confined in the local field of ions.* ChemPhysChem, 2014. **15**(18): p. 4077-4086.

61. Muncan, J. and R. Tsenkova, *Aquaphotomics—From innovative knowledge to integrative platform in science and technology.* Molecules, 2019. **24**(15): p. 2742.

62. Wypych, G., *Handbook of solvents*. 2001, Toronto; New York: ChemTec.

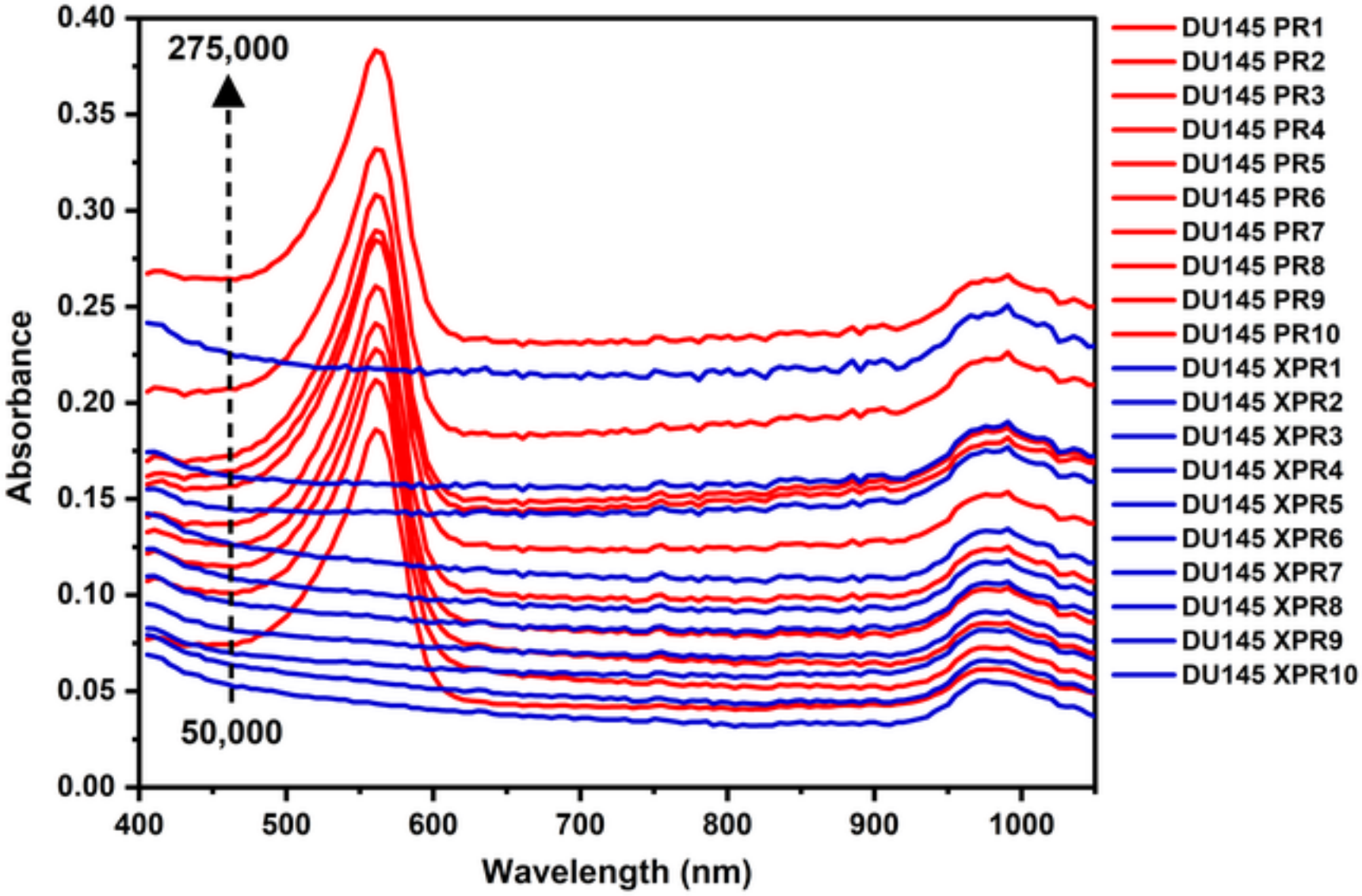
63. Day, C.-P., et al., *Recognition of observer effect is required for rigor and reproducibility of preclinical animal studies.* Cancer Cell, 2022. **40**(3): p. 231-232.

Highlight

- HeLa, DU145, and L929 cells were grown in culture medium with and without phenol red dye.
- Number of the cultured cells for L929, DU145, and HeLa was ranging from 50,000 to 275,000.
- Absorbance spectra were recorded for each cell line at each number of cells.
- Quantification of cells number using MLR and PLSR models in the NIR region produced high prediction accuracy.
- The presence of phenol red in the culture medium improved the prediction accuracy in the case of cancerous cells.

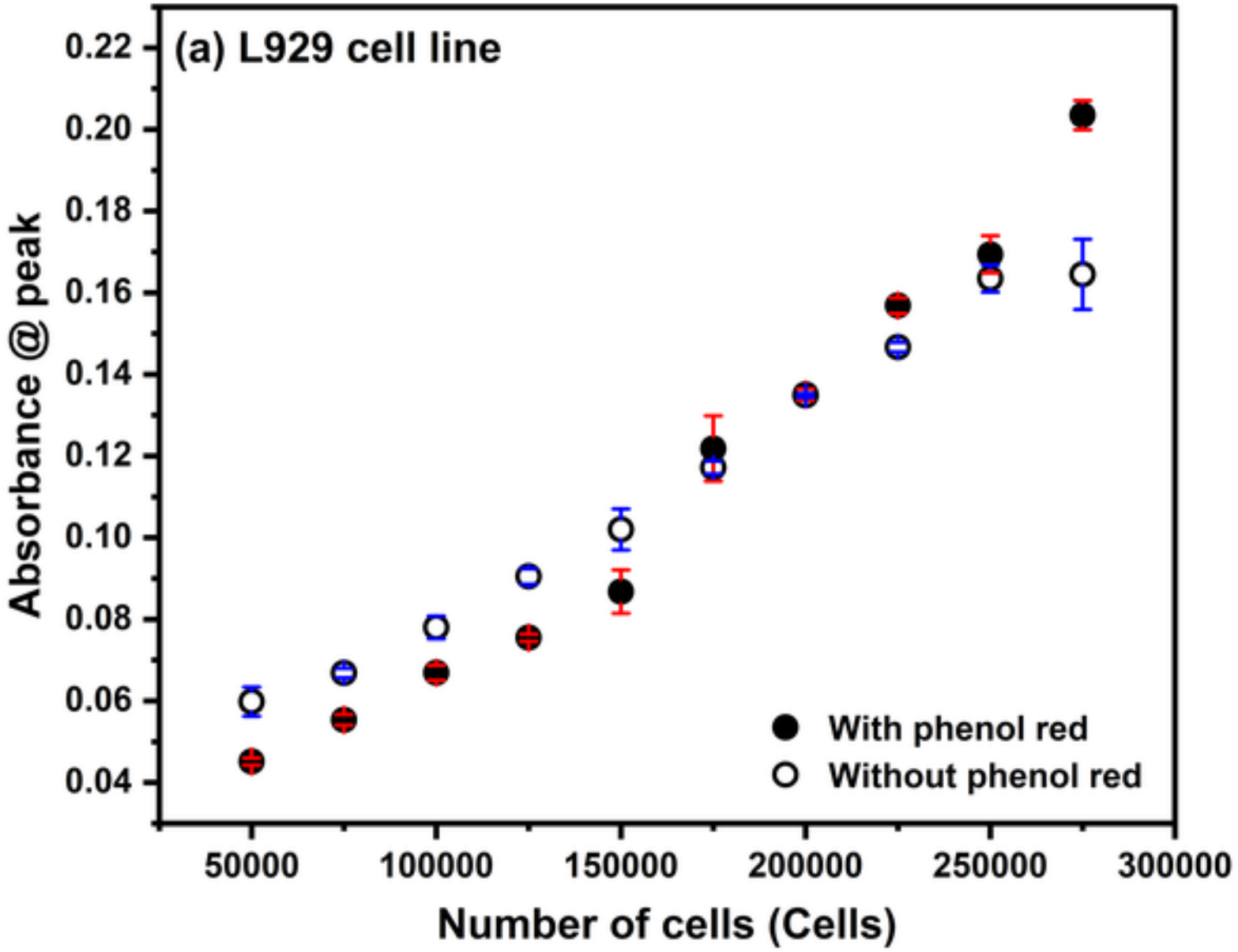
Figure

Click here to access/download;Figure;Figure 1.jpg



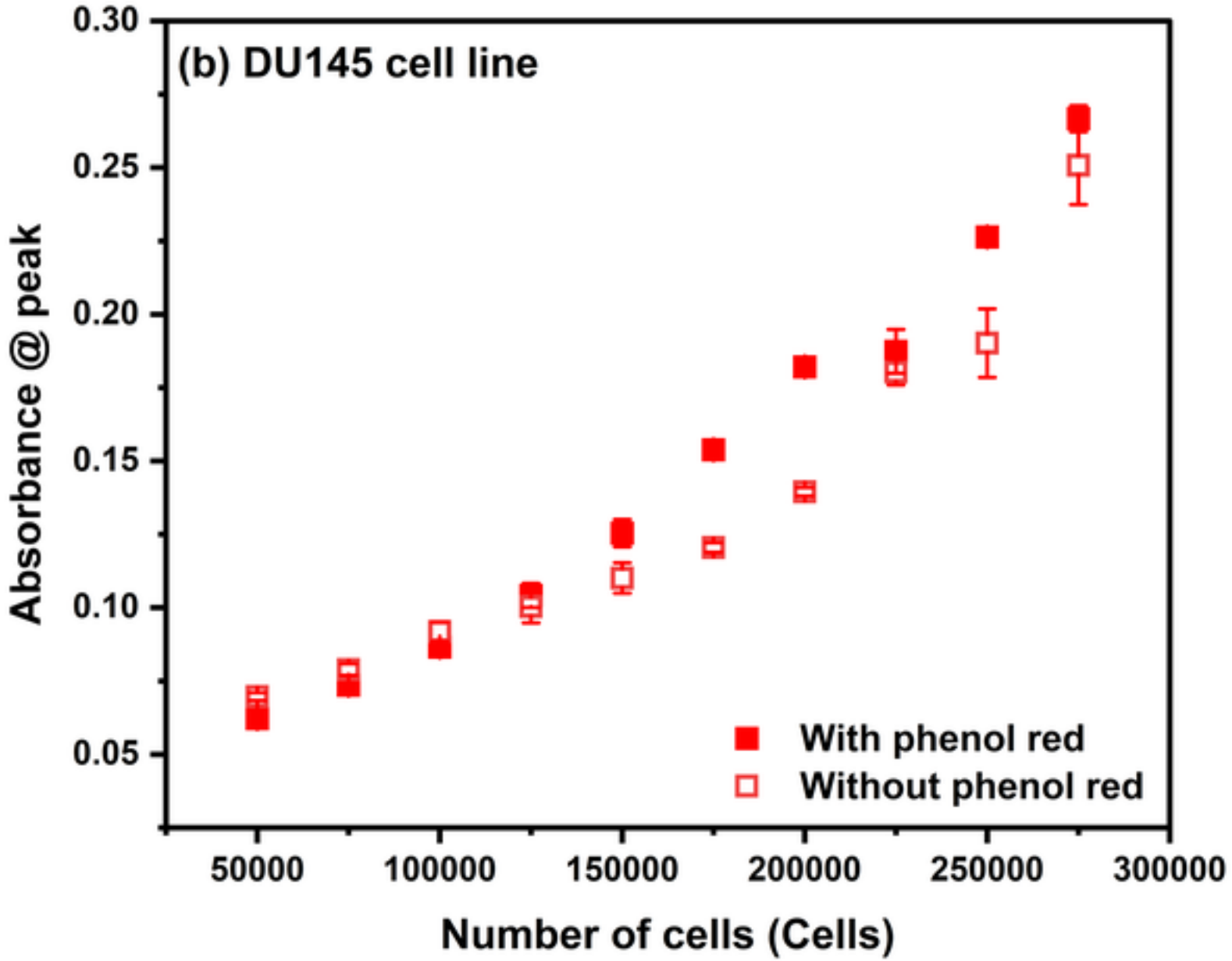
Figure

Click here to access/download;Figure;Figure 2a.jpg



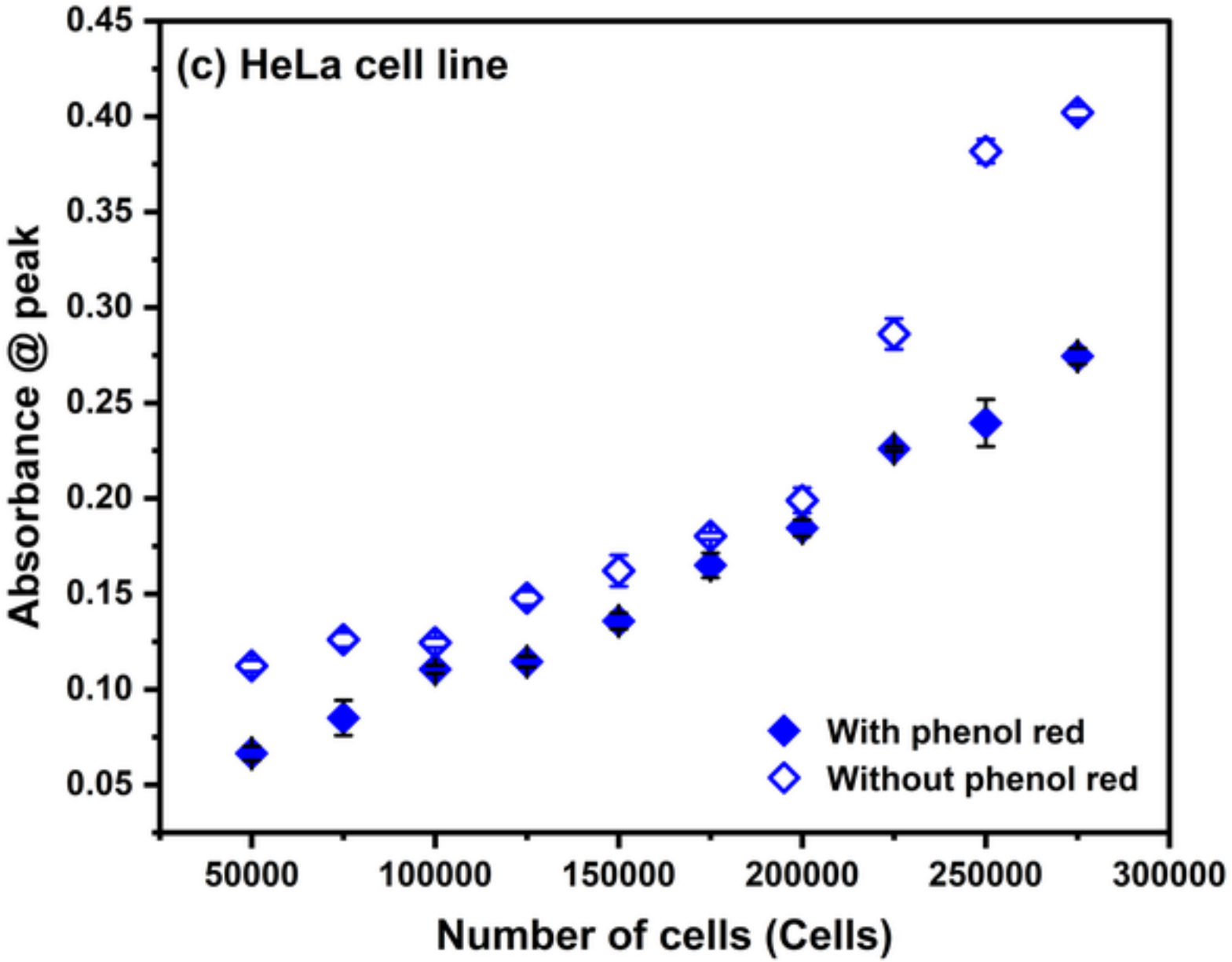
Figure

Click here to access/download;Figure;Figure 2b.jpg



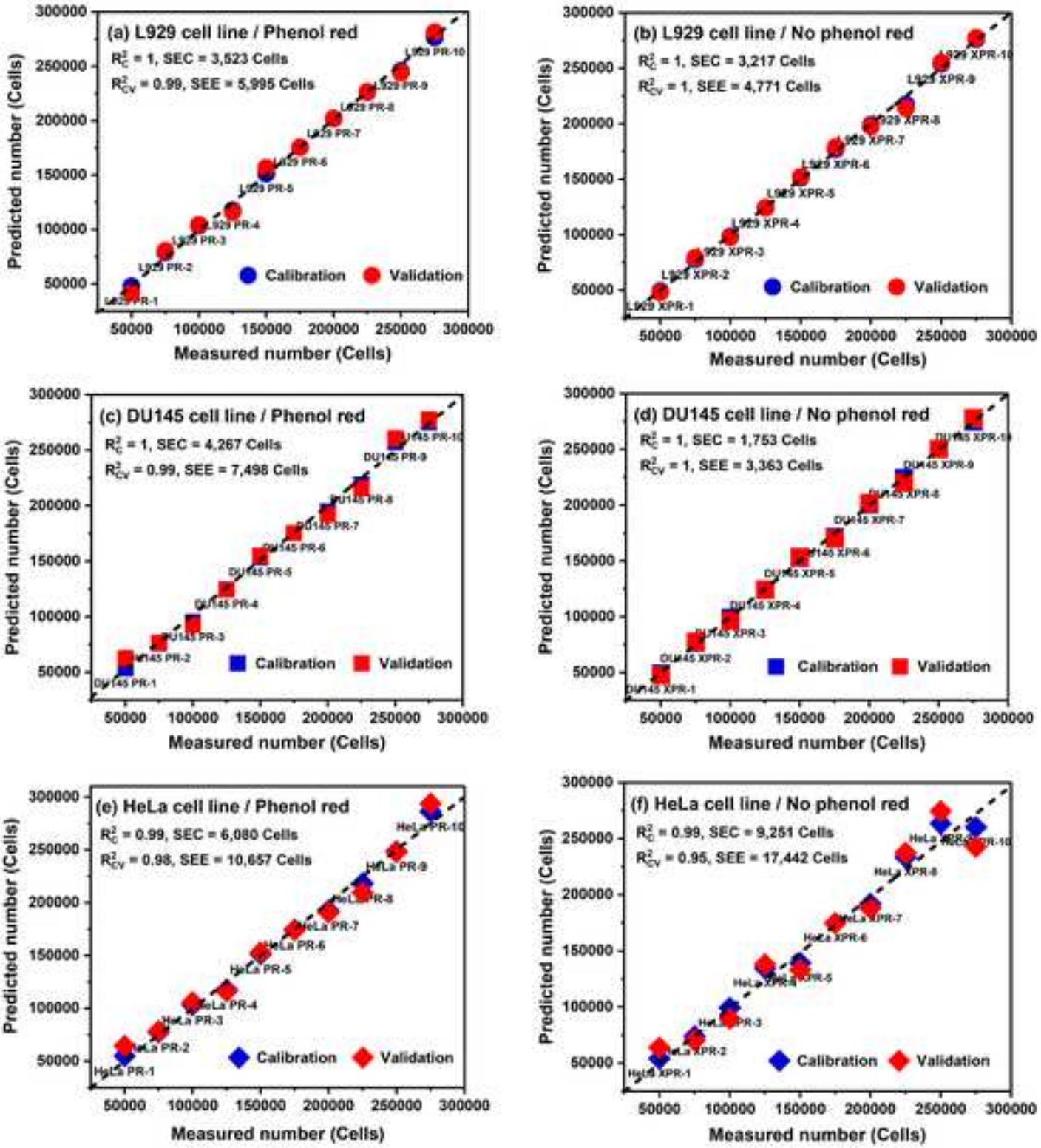
Figure

Click here to access/download;Figure;Figure 2c.jpg



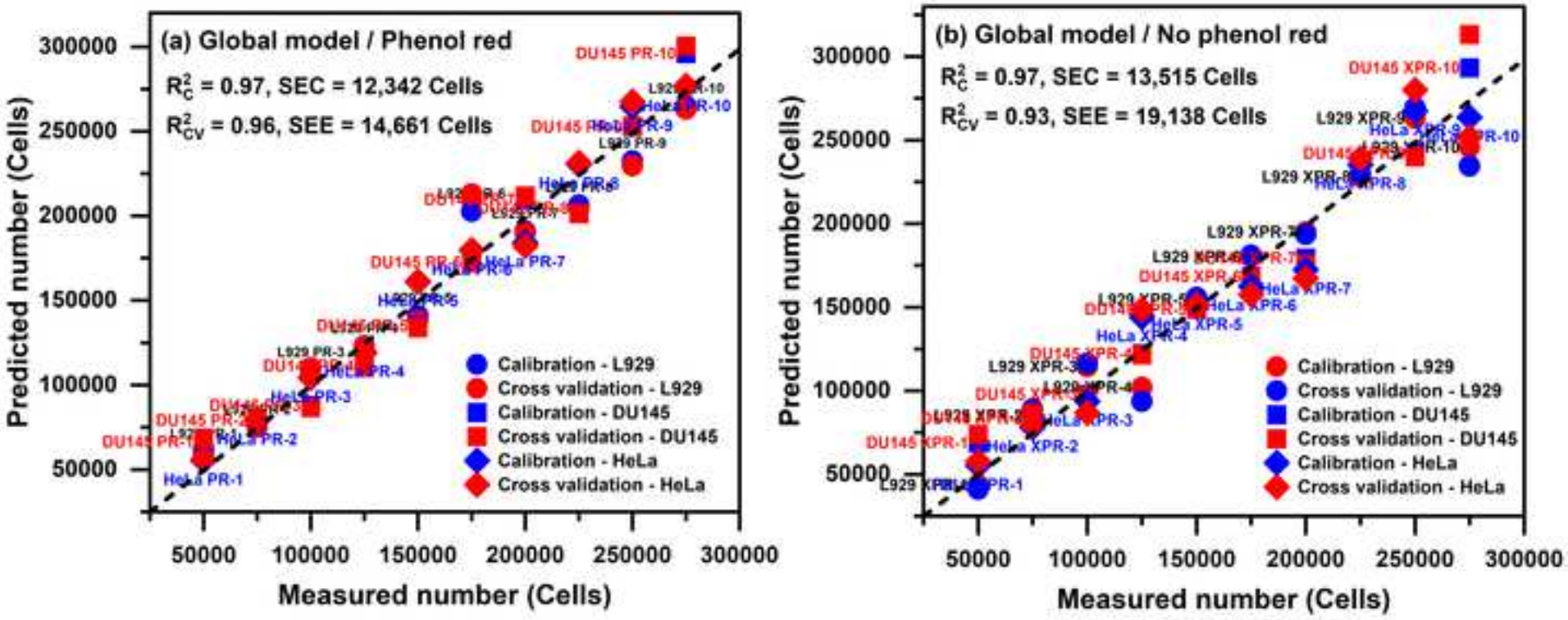
Figure

Click here to access/download;Figure;Figure 3.jpg



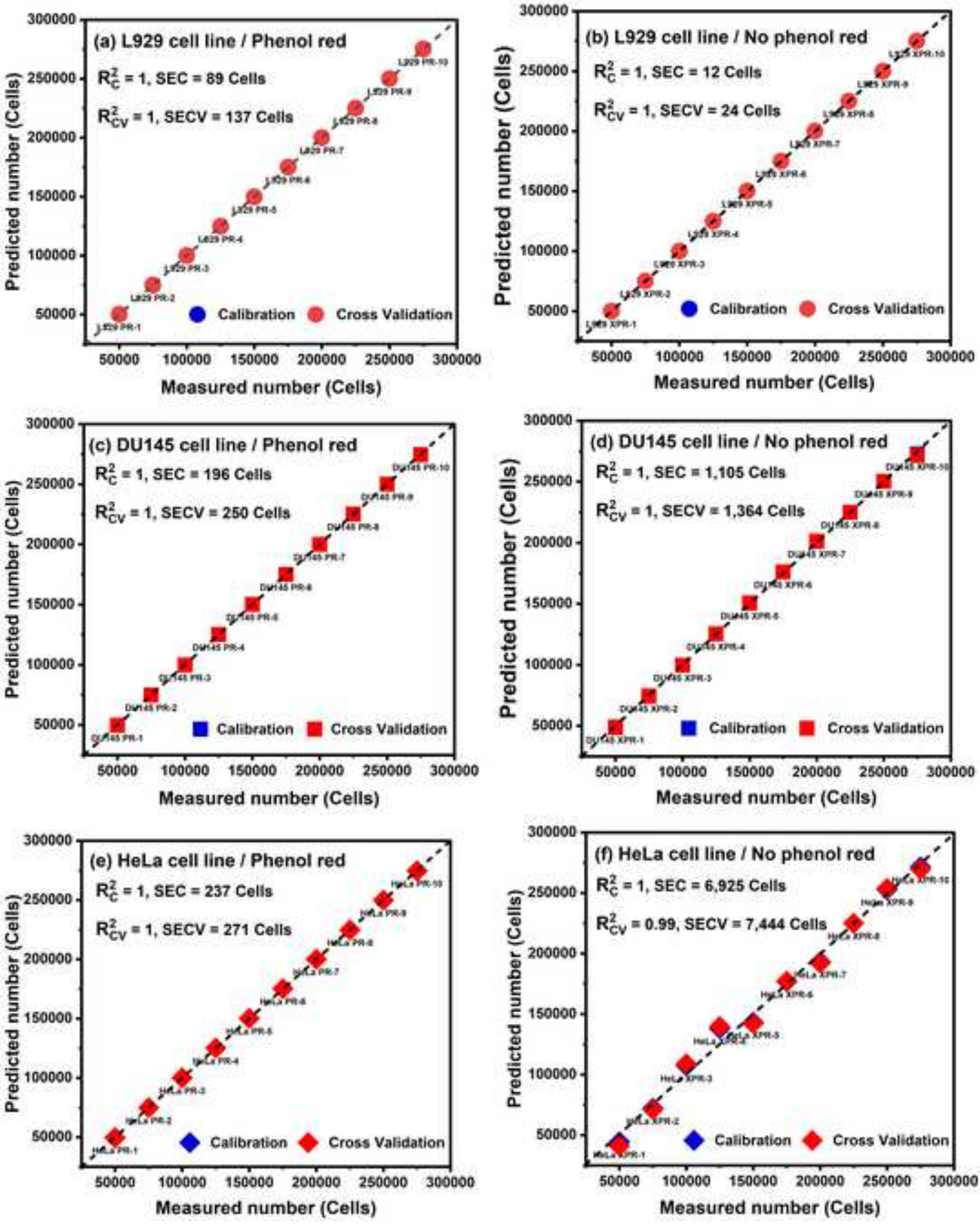
Figure

Click here to access/download;Figure;Figure 4.jpg



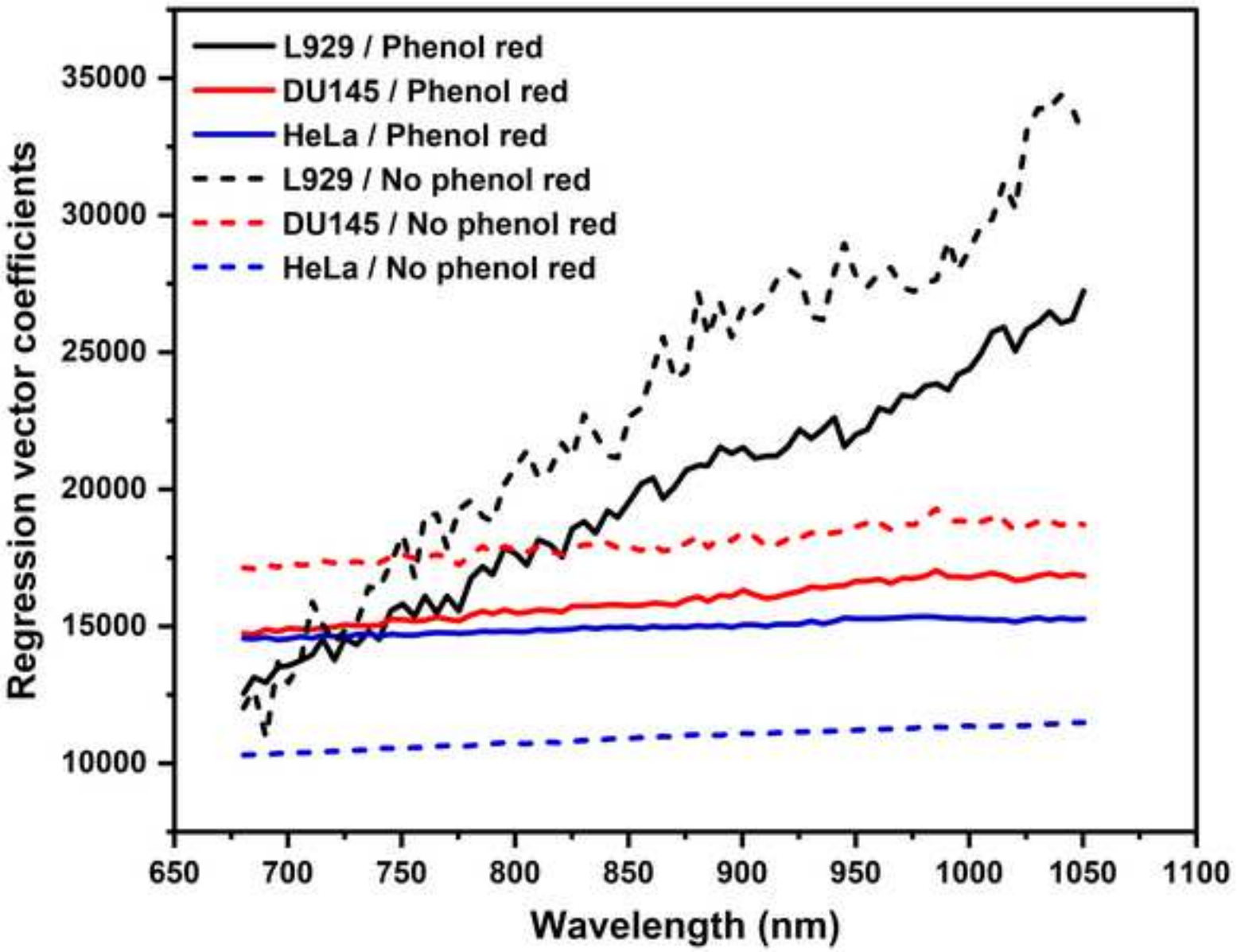
Figure

Click here to access/download;Figure;Figure 5.jpg



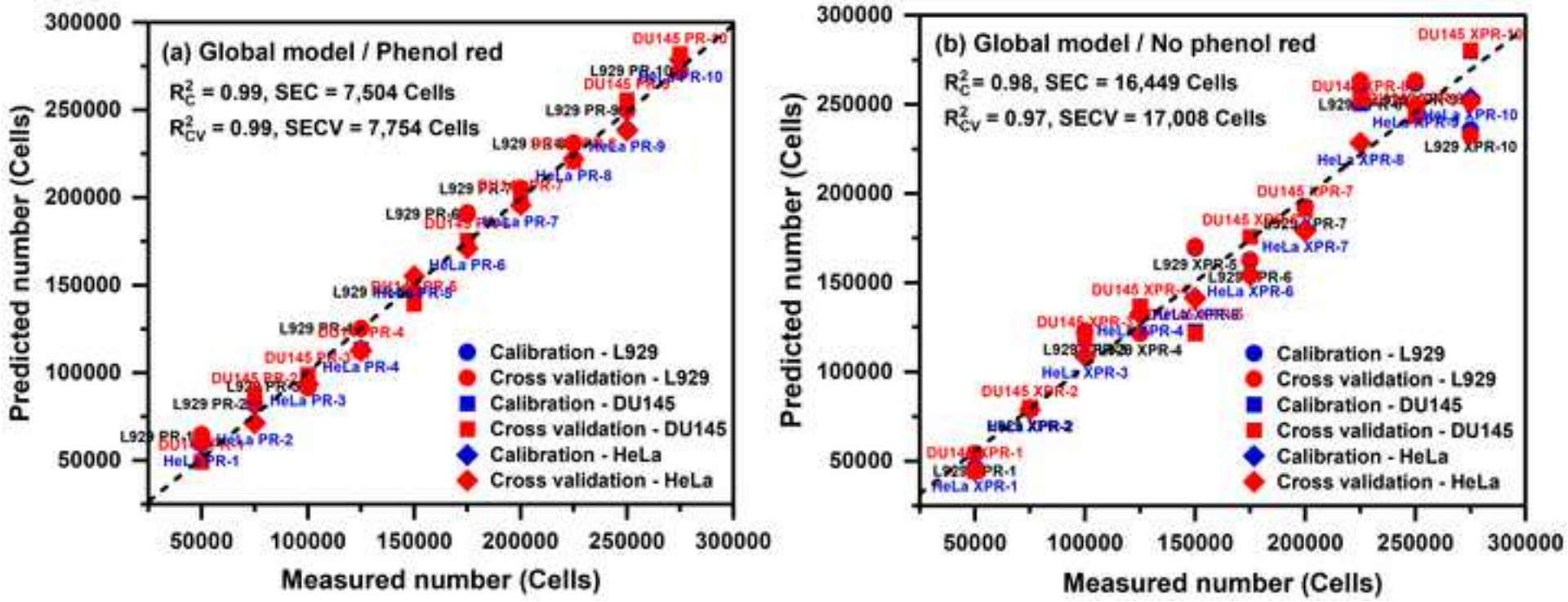
Figure

Click here to access/download;Figure;Figure 6.jpg



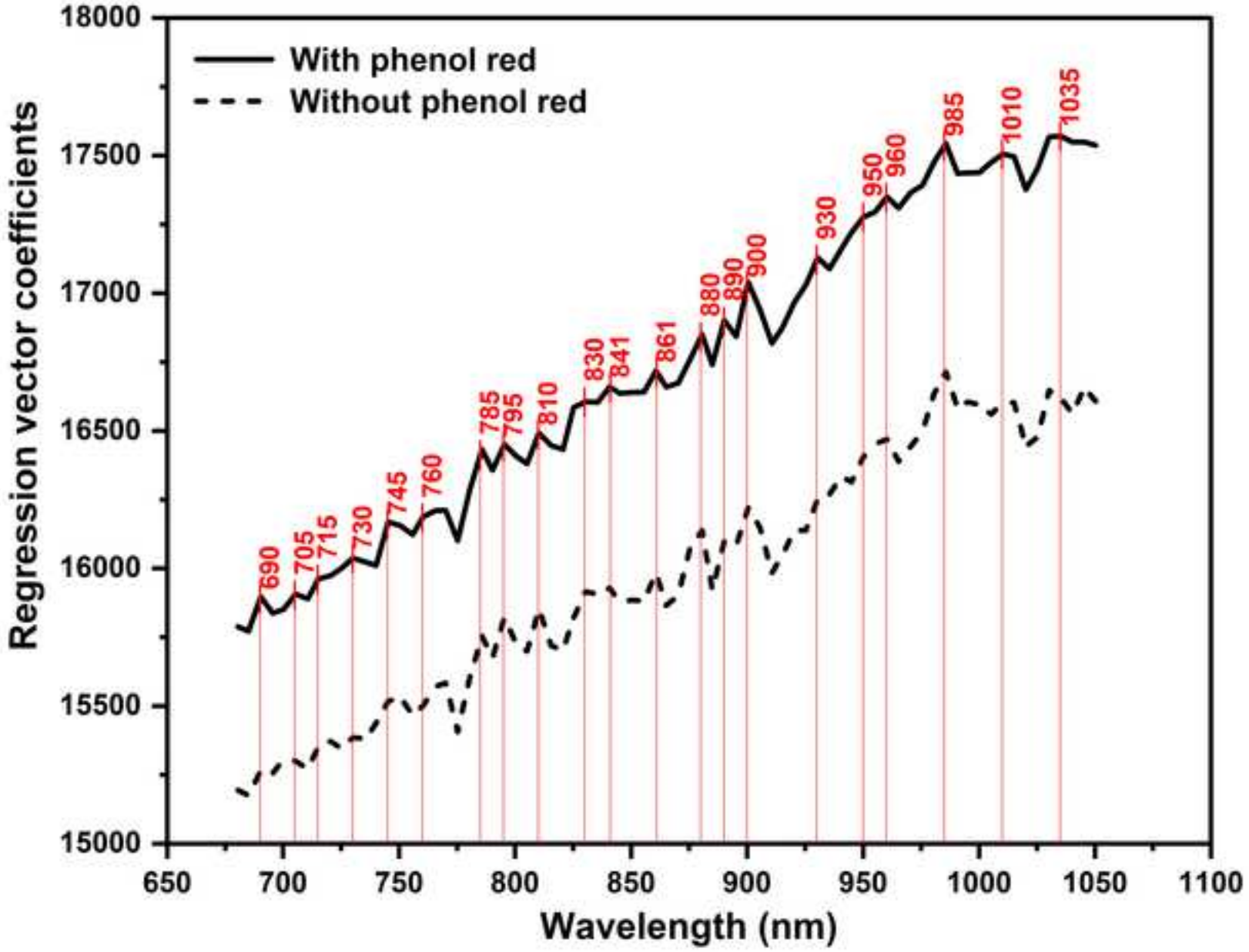
Figure

Click here to access/download;Figure;Figure 7.jpg



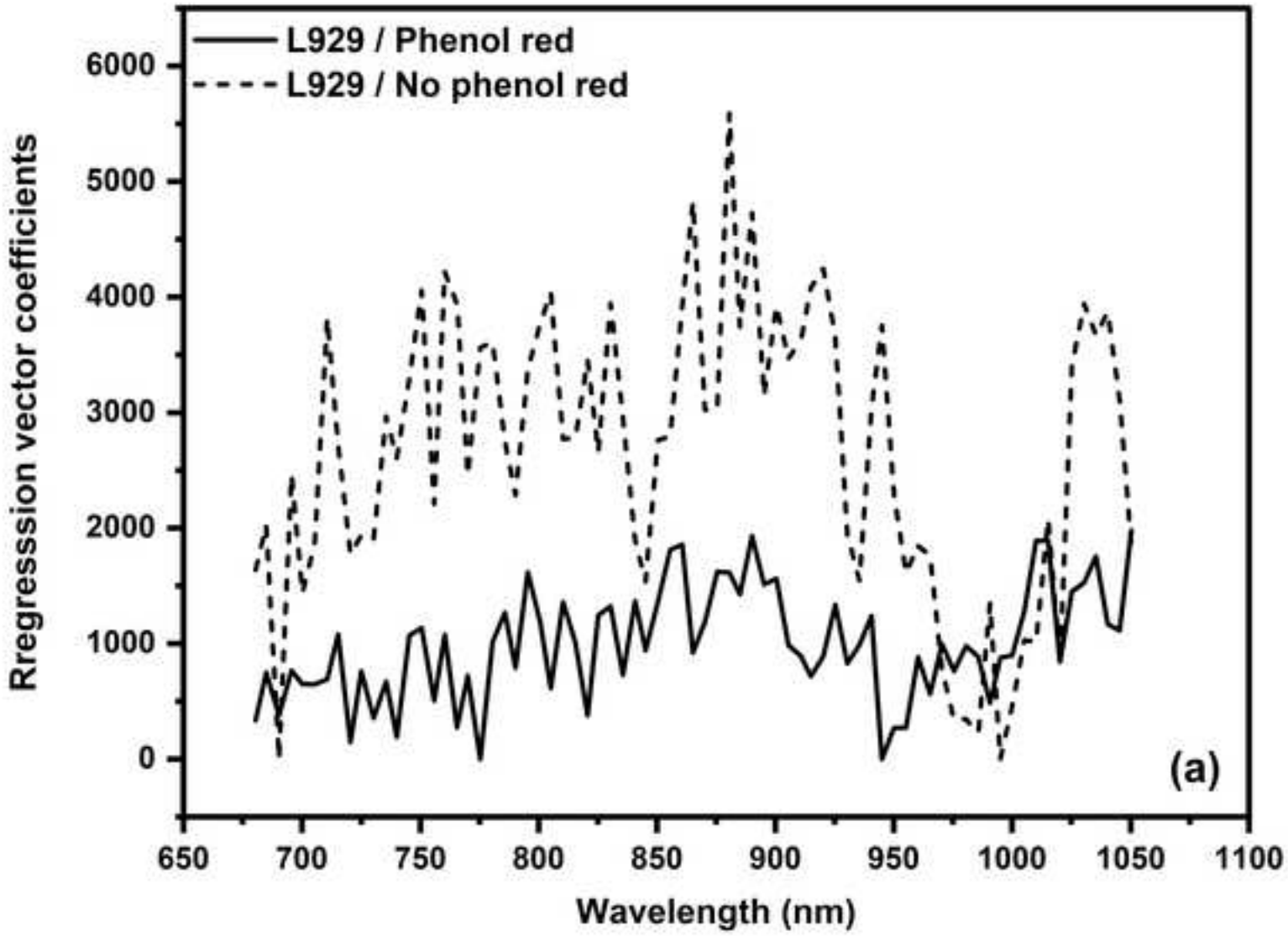
Figure

Click here to access/download;Figure;Figure 8.jpg



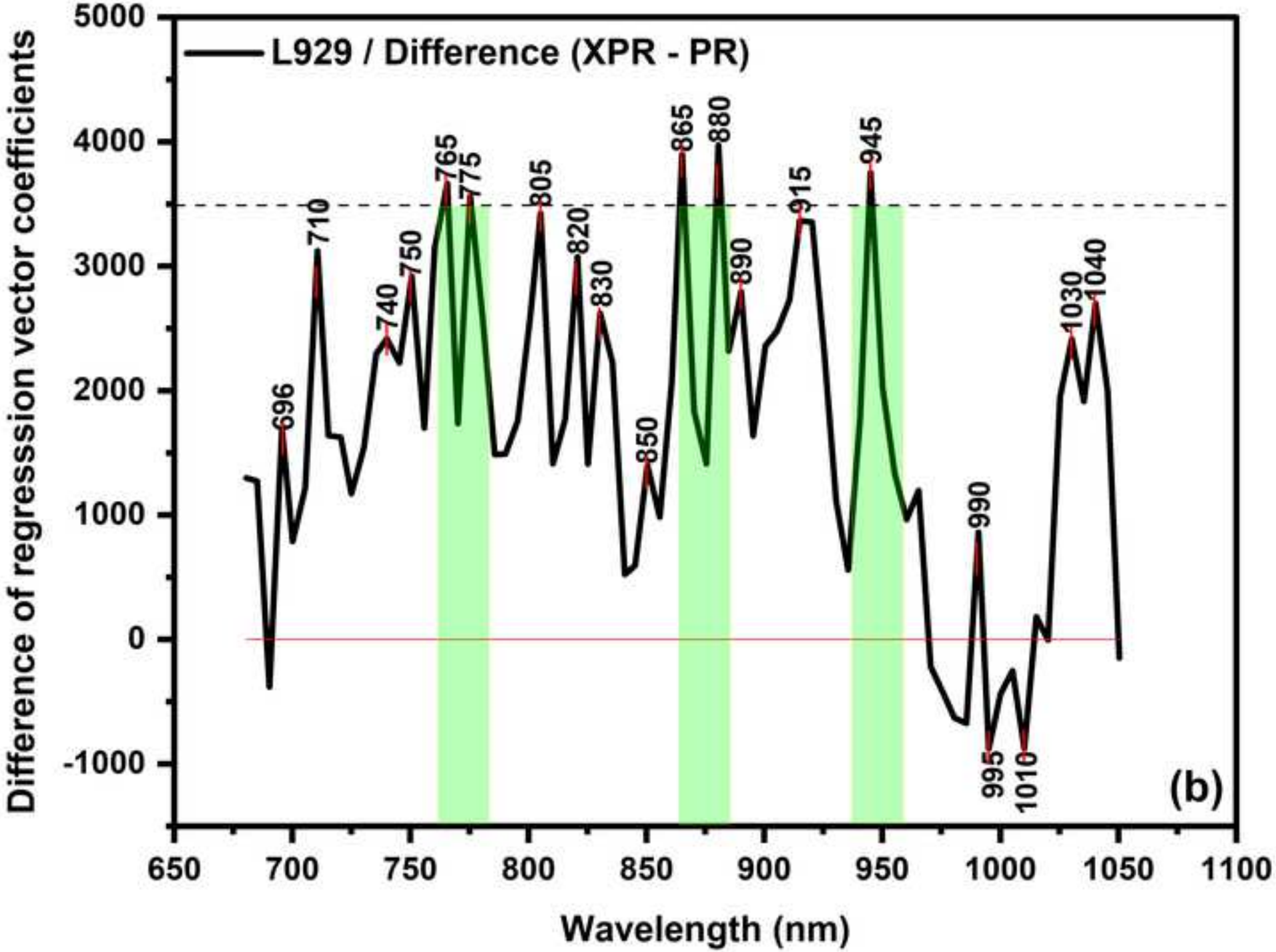
Figure

Click here to access/download;Figure;Figure 9a.jpg



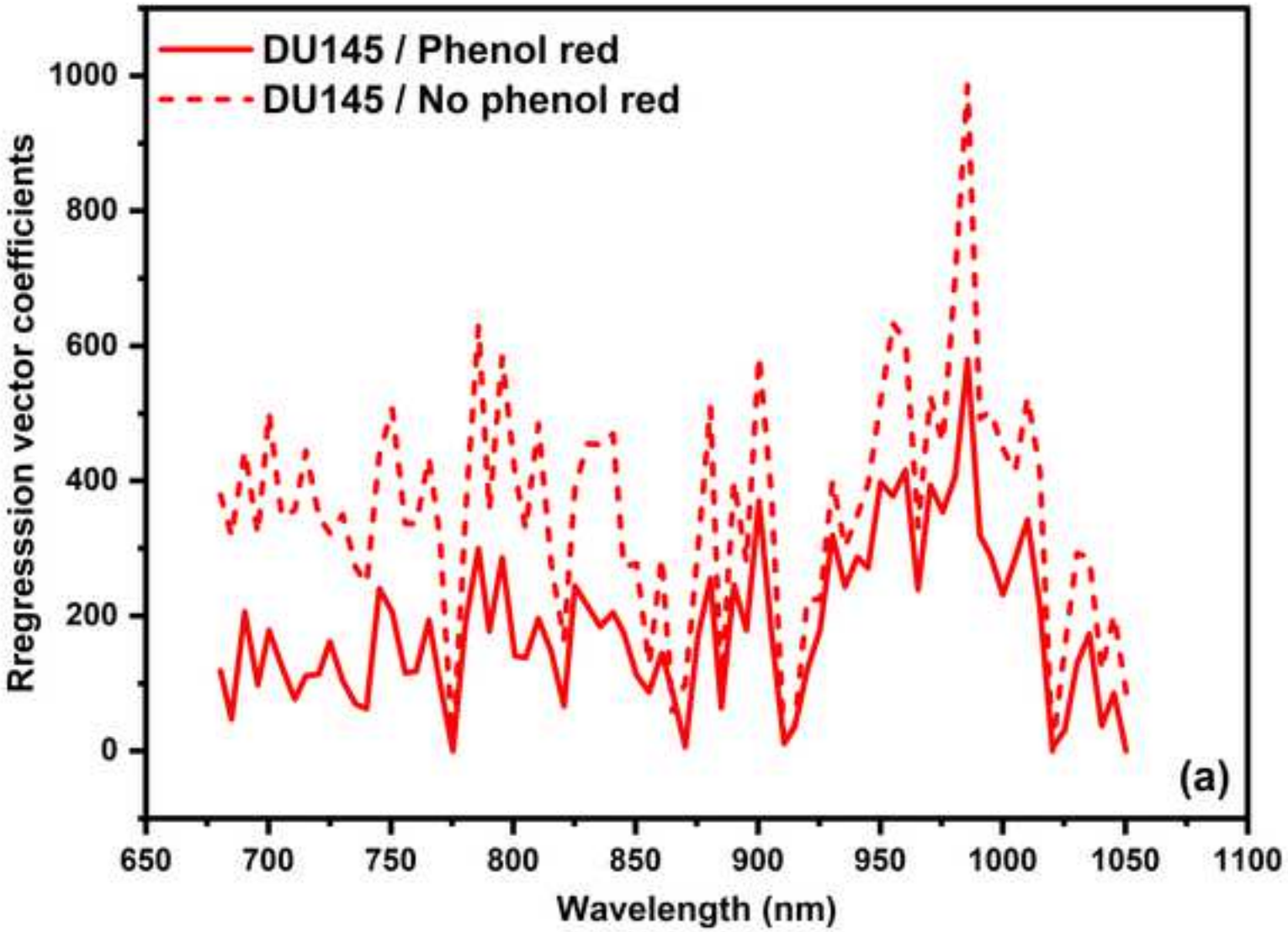
Figure

Click here to access/download;Figure;Figure 9b.jpg



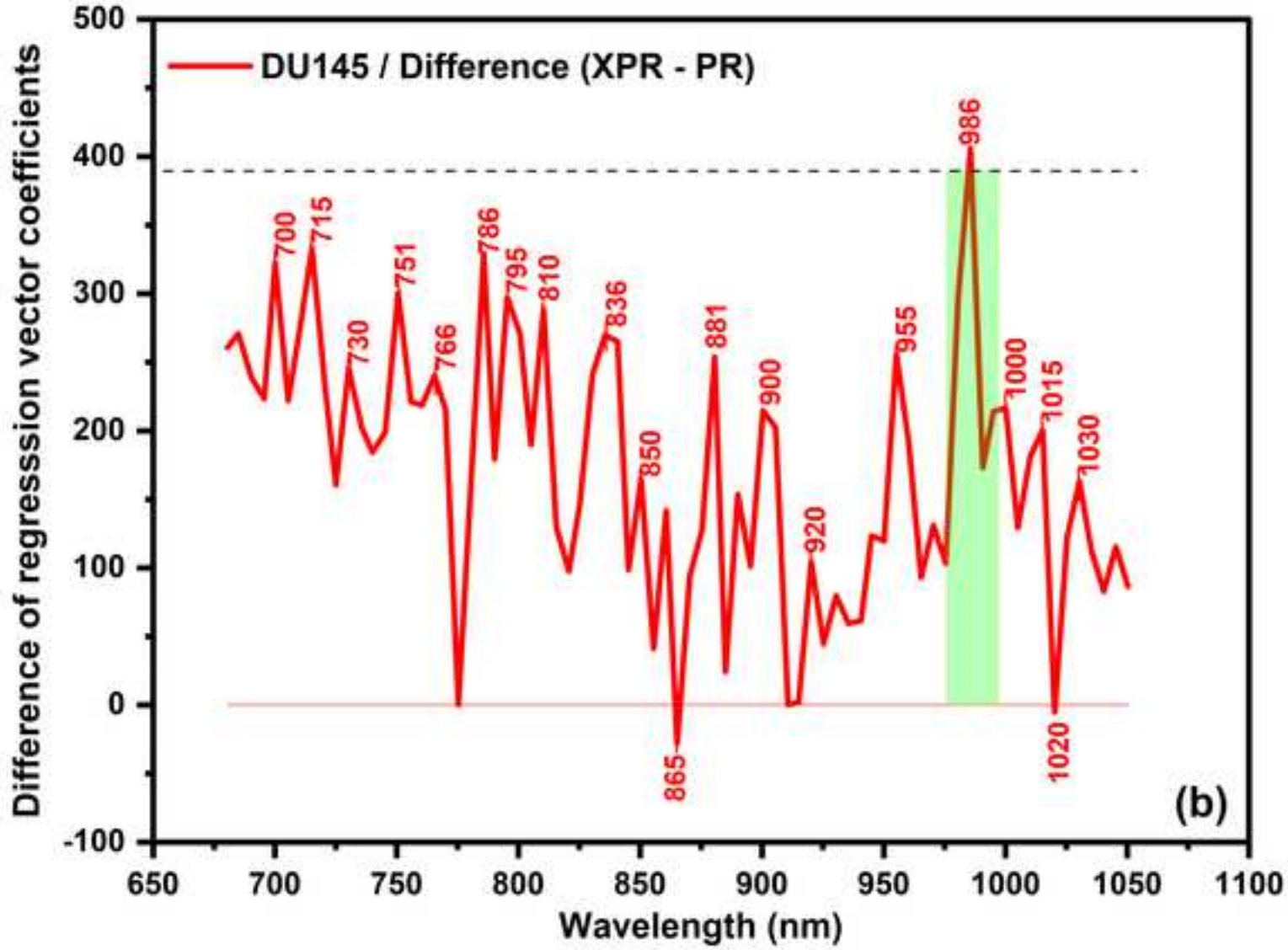
Figure

Click here to access/download;Figure;Figure 10a.jpg



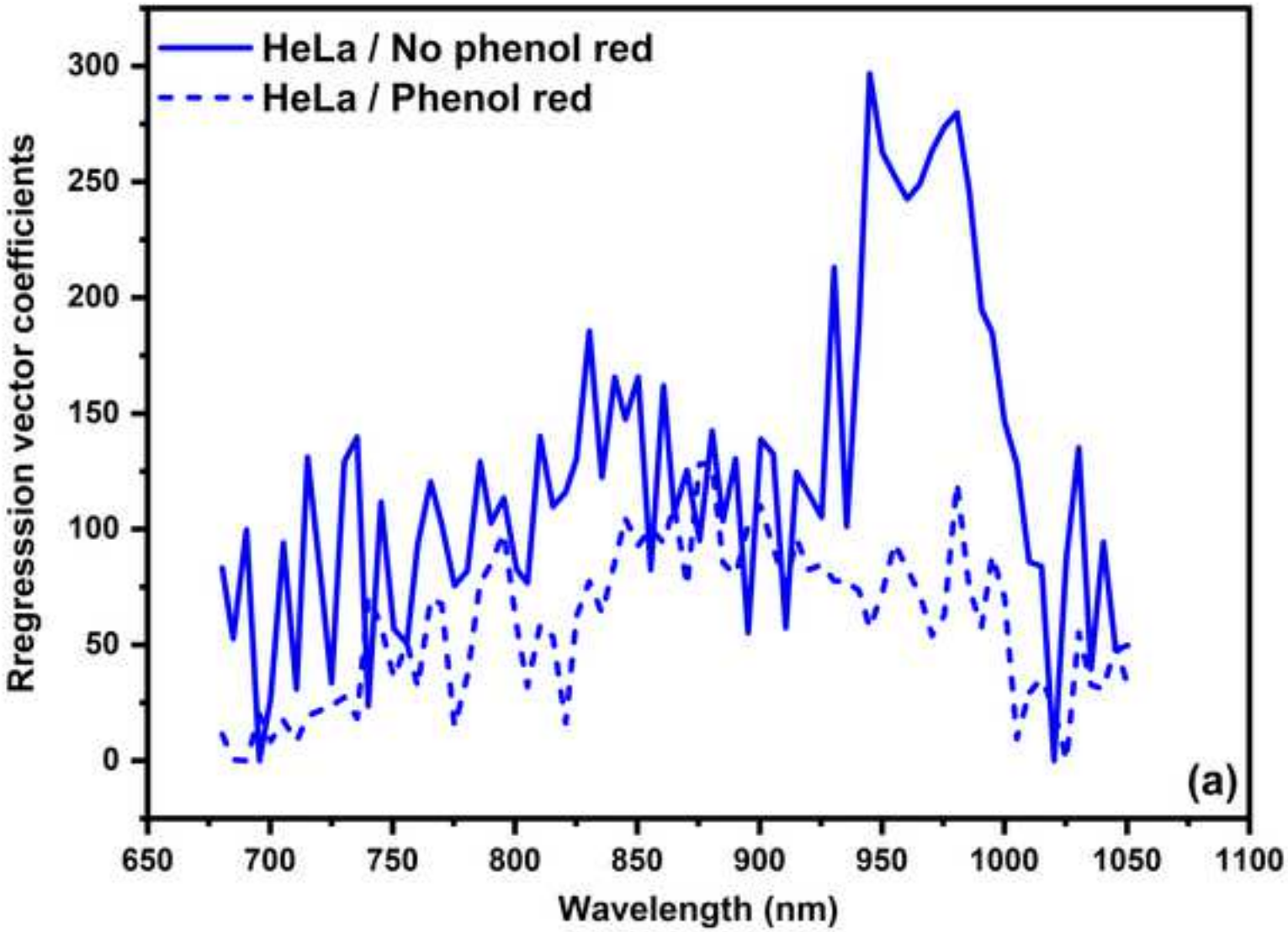
Figure

Click here to access/download;Figure;Figure 10b.jpg



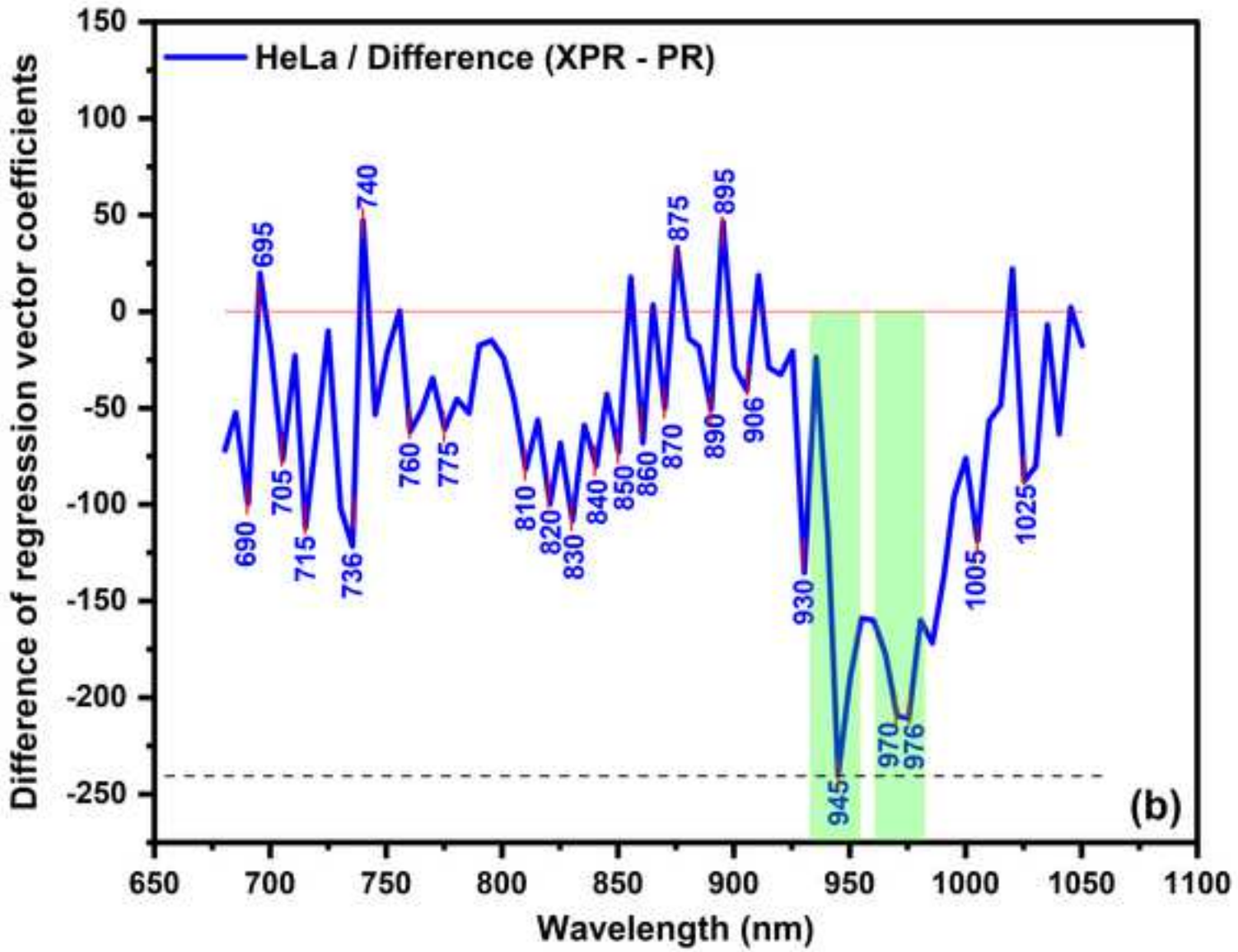
Figure

Click here to access/download;Figure;Figure 11a.jpg



Figure

Click here to access/download;Figure;Figure 11b.jpg



With phenol red											
L929				**DU145**				**HeLa**			
SWs (nm)	R^2_{CV}	SEE (Cells)	α	SWs (nm)	R^2_{CV}	SEE (Cells)	α	SWs (nm)	R^2_{CV}	SEE (Cells)	α
775, 841, 955	0.99	5,995	2.22E-08	700, 801, 976	0.99	7,498	7.01E-08	700, 801, 976	0.98	10,657	5.86E-07
Without phenol red											
L929				**DU145**				**HeLa**			
SWs (nm)	R^2_{CV}	SEE (Cells)	α	SWs (nm)	R^2_{CV}	SEE (Cells)	α	SWs (nm)	R^2_{CV}	SEE (Cells)	α
736, 941, 1015	1	4,771	1.29E-08	736, 775, 805, 906	1	3,363	2.33E-08	810, 861, 960, 986	0.95	17,442	9.44E-05

With phenol red			
SWs (nm)	R^2_{CV}	SEE (Cells)	α
736, 775, 805	0.96	14,661	3.47E-20
Without phenol red			
SWs (nm)	R^2_{CV}	SEE (Cells)	α
745, 810, 861, 900, 960, 986, 1010	0.93	19,138	1.22E-14

Case	Pre-processing method	L929			DU145			HeLa		
		LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)
With phenol red	None	5	0.99	12,059	3	0.99	11,685	4	0.99	8,064
	Smoothing (11 points)	2	0.99	12,517	4	0.99	9,348	3	0.99	11,275
	SNV	4	0.89	34,022	6	0.86	37,709	5	0.79	44,570
	MSC	5	0.58	73,835	2	0.83	41,045	5	0.8	43,929
	1st Derivative (15 points)	7	0.69	55,121	6	0.93	27,399	6	0.89	34,213
	2nd Derivative (15 points)	3	0.76	46,852	3	0.96	21,256	2	0.78	45,231
	OSC	**2**	**1**	**137**	**1**	**1**	**250**	**1**	**1**	**271**
Case	**Pre-processing method**	**L929**			**DU145**			**HeLa**		
		LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)
Without phenol red	None	2	0.99	5,570	5	0.96	23,077	2	0.96	21,032
	Smoothing (21 points)	2	0.99	5,625	4	0.97	19,351	2	0.96	20,993
	SNV	2	0.78	45,562	3	0.96	18,939	1	0.78	46,253
	MSC	2	0.78	45,525	3	0.96	19,008	1	0.76	47,500
	1st Derivative (15 points)	2	0.62	60,557	3	0.97	18,751	1	0.78	44,847
	2nd Derivative (15 points)	1	0.68	54,716	1	0.92	27,577	3	0.74	48,440
	OSC	**3**	**1**	**24**	**1**	**1**	**1,364**	**1**	**0.99**	**7,444**

Case	Pre-processing method	LVs	R^2_{CV}	SECV (Cells)
With phenol red	None	5	0.98	13,007
	Smoothing	7	0.98	13,694
	SNV	3	0.73	50,343
	MSC	3	0.62	60,740
	1st Derivative	5	0.88	34,176
	2nd Derivative	7	0.75	49,802
	OSC	**1**	**0.99**	**7,754**
Case	**Pre-processing method**	**LVs**	R^2_{CV}	**SECV (Cells)**
Without phenol red	None	5	0.93	26,577
	Smoothing	5	0.94	25,447
	SNV	3	0.79	43,826
	MSC	3	0.8	43,606
	1st Derivative	5	0.88	34,380
	2nd Derivative	2	0.82	41,279
	OSC	**1**	**0.97**	**17,008**

Band position in 2nd overtone of water (nm)	**Calculated position in 1st overtone of water (1300 - 1600 nm)**	**Assignment / WAMACs [60-62]**
915	1372.5	C3 – $\nu_1 + \nu_3$: Water symmetrical stretching vibration and water asymmetric stretching vibration
930	1395	C5 – Water molecules confined in the local field of ions
945	1417.5	C5 – Free water molecules, S_0
955	1432.5	C6 – S1: Water molecules with 1 hydrogen bond
970	1455	C8 – Water solvation shell, $OH^-(H_2O)_{4,5}$
976	1464	C9 – Water molecules with 2 hydrogen bonds, S_2
986	1479	C10 – Water molecules with 3 hydrogen bonds, S_3

Declaration of Interest Statement

The authors declare that they have no conflict of interest.

Corresponding Author

Ahmad Fairuz Omar

Authors' Contribution:

The authors confirm contribution to the paper as follows: study conception and design: Ahmad Fairuz Omar; experimental work: Suhainah Sudik; manuscript writing and preparation: Muna E. Raypah; analysis and interpretation of results: Muna E. Raypah and Jelena Muncan; laboratory training and assistance: Mohd Hafiz Mail and Azman Seeni; approved the final version of the manuscript: Roumiana Tsenkova; supervision and grant: Ahmad Fairuz Omar.

Implication of Phenol Red in Quantification of Cultured Cancerous Cells using Near-Infrared Spectroscopy and Aquaphotomics

Muna E. Raypah [1*], Jelena Muncan [2**], Suhainah Sudik [1], Ahmad Fairuz Omar [1***], Mohd Hafiz Mail [3], Roumiana Tsenkova[2], and Azman Seeni [4]

[1] School of Physics, Universiti Sains Malaysia, 11800, Pulau Penang, Malaysia
[2] Aquaphotomics Research Department, Faculty of Agriculture, Kobe University, Kobe, Japan
[3] Malaysian Institute of Pharmaceuticals and Nutraceuticals, National Institute of Biotechnology Malaysia, Ministry of Energy, Science, Technology, Environment and Climate Change, 11700 Penang, Malaysia
[4] Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200, Pulau Penang, Malaysia
*Corresponding authors' e-mails: fairuz_omar@usm.my, muna_ezzi@usm.my, , and jmuncan@people.kobe-u.ac.jp

Abstract

This study aims to evaluate the ability of absorbance spectra in the near-infrared (NIR) region to predict the number of cells for different cell lines. The cancerous cell lines were human cervix adenocarcinoma (HeLa) and human prostate carcinoma (DU145), and L929 was a normal mouse skin fibroblast. The number of cells varied from 50,000 to 275,000 with an interval of 25,000 for each cell line. Vis-NIR absorbance spectra (400-1100 nm) at each number of cells (50,000-275,000) for L929, DU145, and HeLa cell lines cultured in media with and without phenol red were recorded. Multiple linear regression (MLR) and partial least squares regression (PLSR) models were developed in the NIR region (680-1050 nm) to quantify the number of cells for the three cell lines. The outcomes showed that the quantification analysis of the number of cells using MLR and PLSR models produced high prediction accuracy with $R^2 \geq 93\%$. The best results were obtained using PLSR for the preprocessed spectra using an orthogonal signal correction method. It was found that the presence of phenol red in the culture medium improved the prediction accuracy in the case of HeLa (SECV = 271 cells) and DU145 (SECV = 250 cells) cell lines, but the accuracy was higher when phenol red was not present for the L929 cell line (SECV = 24 cells). In addition, the existence of phenol red boosted the accuracy when the global PLSR model was built, irrespective of the cell type (SECV = 7,754 cells). The effect of phenol red was explained in the terms of its impact on the water molecular structure of the cells' culture medium which influences the light scattering.

Keywords: Cancerous Cells; Phenol Red; Near-infrared Spectroscopy; Quantification; MLR; PLSR.

1. Introduction

Cancer is a complex disease, typified by reprogrammed signaling, cellular physiological alterations, and uninhibited cell growth. A characteristic trait of cancer cells is their metabolic reprogramming, which facilitates fast cellular reproduction, migration, and alteration of their microenvironment, ultimately enabling metastasis [1]. Early detection and early-stage surgical and chemotherapeutic interventions and therapies help inhibit cancer growth. Nonetheless, cancer becomes untreatable or terminal when it metastasizes. The gradual growth of a tumor allows for the early detection of a malignancy (or pre-malignancy) which lowers the prevalence of late-stage cancer [2]. The early detection of tumors is crucial for effective cancer treatment and for elucidating tumorigenesis. Expanding the range of recognized reliable biomarkers may aid the discrimination of benign cancers from malignant ones by clinicians, thus decreasing the incidence of needless biopsies, and enhancing the effectiveness of cancer treatments.

The recent advancement in point-of-care diagnostics and precision medicine has prompted the demand for inexpensive and accurate cell counting technologies [3]. The proliferation of tumor cells is determined primarily by three key parameters: the progression period for reproducing cells, a part of proliferating cells, and the quantity of unprompted cell damage [4]. Growth rates of solid tumors can be determined without difficulty along the linear portion of the Gompertzian growth curve. Visible and detectable tumors typically hold between 10^6 to 10^9 cells [5, 6]. The current detection threshold for solid tumors proliferating as a single mass is in the region of 10^9 cells ($1\ \mathrm{g} = 1\ cm^3$) [5]. Numerous cases of different solid tumors have shown linearity of growth on a logarithmic scale throughout the early detection stage.

Globally, treatment centers and hospitals extensively utilize cell counting techniques for the determination of a patient's health [3]. Clinically, a tumor with a volume of 1 cm^3 ($\approx 10^9$ cells)

is considered rather small, and at this size, the tumor may induce the initial symptoms and can be perceptible on palpation, or with the use of laboratory diagnostic examinations. The most frequent error is to consider a small tumor as an ‘‘early’’ tumor. At the stage when it becomes detectable, a malignant tumor has existed for a minimum of half of its lifespan [4]. The significance of early tumor detection for the diagnosis of the patient has been overly estimated since tumors cannot be clinically detected early. Cancer may have spread long before the primary tumor can be diagnosed. The precise determination of the number of cells in culture is crucial for experimental reproducibility and standardization [7]. Cell counts are vital for the evaluation of cell health and proliferation rate, assessment of immortalization or alteration, seeding cells for successive experiments, transfection or infection, and preparation for cell-based assays. The cell counts must be precise, reliable, and rapid, especially for the quantitative determination of cellular responses [8]. There is also an immense need for the precise quantification of biomarkers for cancers [9], particularly at the early recognition of a miniature volume of tumors [10].

Quantitative techniques are advantageous due to the insight they offer into understanding the biology/physiology of tumor cells beyond diagnosis. The method that is most commonly used for direct cell counting is trypan blue staining using a hemacytometer. This approach allows counting the number of cells manually under a microscope. However, the hemacytometer-based cell enumeration can be very biased, subject to errors due to device misuse, tiresome, and time-consuming. Coulter counter (automated counter) was the first electronic cell counter which was widely accepted as an alternative to manual counting. Automated cell counters based on the Coulter counter principle, optical flow cytometry, and image-based cytometry have been established and commercially available. These devices can detect cells using integrated software to describe cells in terms of their fluorescent intensity and size in a few seconds. It was reported a

high precision and accuracy were obtained in the case of using the optical flow cytometer in comparison to manual and Coulter counting methods [11]. Currently, in research laboratories, images of cells obtained using microscopy are also commonly examined and quantified using open-source software [12]. Cell profiler, ImageJ, and other such open-source software are routinely employed by cell biologists for cell enumeration from images [13-15]. This can be achieved by performing object identification through segmentation, thresholding, recognition, and division of clumped cells and other processing steps. At present, most of the microfluidics-based cell enumerating systems available in the market utilize a microfluidic chip or cartridge combined with a traditional cytometry system [16]. Such designs offer ease and flexibility since they include a fluidics system with microchannels for sample flow. However, the microfluidic devices still need more improvement in sample preparation and detection techniques within a lab-on-a-chip microfluidic platform [3]. Therefore, there is increasing demand for an objective cell quantification technique that produces an accurate and reliable number of cells.

The quantification of predictive biomarkers for cancer is significant, particularly at the early stage of cancer when the number of cells is small. In addition, the quantification of cells is essentially significant in both fields of pathology and clinical studies [17] and is considered a vital approach to elucidating the cellular composition, growth, development of diseases, and aging. Recently, the optical spectroscopy technique has unveiled its potential undamaging capability to exhibit distinct spectral attributes that discriminate between tumors and normal cells. The spectroscopy techniques are rapid, simple to use, environmental-friendly, cost-effective, and non-destructive. Thus far, spectroscopy techniques have been extensively utilized for the diagnosis of a range of cancer types such as lung, oral, leukemia, breast, and prostate cancer [18-24]. However, there is inadequate data on the optical spectroscopy properties of cancer cells, particularly for the

quantitative assessment [25]. The potential to attain quantitative characteristics of cells directly from the living system in situ is important. There is a need to comprehensively characterize and establish spectroscopy signatures of cancer at the cell level to better understand the spectral responses. So far, information on cancer cells using the NIR spectroscopy technique is not extensively available, especially for the quantitative assessment of the cultured cells. The supposed "optical window" that occurs in the NIR region 650-1100 nm is appropriate for most of the non-invasive measurements of biological systems [26]. This region is also referred to as the "therapeutic window" for the lower and determinate light absorbance and the abundance of information.

The combination of analytical tools has been proven to be of immense value in a variety of bio-scientific applications. Spectroscopy coupled with chemometrics can offer a rapid method for improving the performance of the cell culture [18] and quantifying the number of cancer cells. Partial least squares regression (PLSR) [27, 28] is one of the most commonly used regression methods for quantification of multivariate spectral data together with multiple linear regression (MLR) and principal component regression (PCR) [29]. It is important to point out that the MLR model is a causative model with clear physical meaning and more explanatory coefficients compared to the coefficients of PLSR. On the other hand, in quantitative studies, the MLR model sometimes cannot provide predictive models which can be considered its main limitation [30]. PLSR is a good method of choice for modelling datasets where the number of variables (wavelengths) surpasses the number of observations (spectra of the samples) and where the predictors are highly correlated [27, 31]; this is considered the main advantage of the PLSR over MLR. In addition, the PLSR model with latent variables that extract the patterns characteristic of descriptor variables provides information about the predictive power of dependent variables [30].

Even though many researchers have worked on cancer cells, very few researchers have reported the quantification of the cancerous cells. Up to now, the spectroscopic analysis combined with chemometrics for quantifying the number of cancerous cultured cells in the NIR region with aid of phenol red has not been reported. This work is an extension of our previous study [25] with an intention to examine the implication of phenol red (a commonly pH indicator in culture media) towards the quantitative analysis of the cancerous cultured cells using the absorbance at the NIR region (700-1100 nm). This research aims to evaluate the accuracy of NIR spectroscopy and aquaphotomics to quantify the number of cultured cancerous cells by developing two regression models. The effect of phenol red was explained in the terms of its impact on the water molecular structure of the cells' culture medium which influences the light scattering. The cancerous cells utilized were cervical (HeLa) and prostate (DU145), while the mouse skin fibroblast (L929) cell line was used as a normal cell. Direct comparisons of the growth of the three cell lines were performed in the culture medium in the presence and devoid of phenol red dye. The MLR was used for a cross-calibration purpose and PLSR was used as another method for quantification of cells using the NIR region from 680 nm to 1050 nm. A comparison was carried out to show how the accuracy of the developed regression models depends on the presence of phenol red in the culture medium. The prediction of cancer cell numbers, particularly at the early stage of cancer, can be potentially applied as a reference feature in cytological analysis.

2. Materials and Methods

2.1. Cell Culture

Three cell lines including mouse skin fibroblast (L929), human prostate carcinoma (DU145), and human cervix adenocarcinoma (HeLa) acquired from American Type Culture Collection

(ATCC), Manassas, VA, USA, and used in this study. The procedure of the cells subculture and the spectroscopic system is similar to those described in our prior reports [25, 32]. The growth of the cells was performed in cell culture media which were Minimum Essential Medium 'MEM' (for L929 and DU145) and Dulbecco's Modified Eagle Medium 'DMEM' (for HeLa) with and without phenol red dye. The media were supplemented with fetal bovine serum 'FBS' (10%), sodium pyruvate (1 mM), and penicillin (100 units/mL)/streptomycin (100 µg/mL). The cell culture was preserved at 37 ℃ in an incubator with a 5% of CO_2 atmosphere. The cells viability test was carried out using a trypan blue dye exclusion technique by quantifying the number of viable cells using a hemocytometer. The cell lines were seeded in a 6-well cell culture plate with a total number of cells ranging from 50,000 to 275,000 with a step of 25,000 per well.

2.2. Spectra Acquisition

The absorbance spectra of each cell line cultured in a medium with and without phenol red were recorded using the Vis-NIR spectroscopy technique. The spectroscopy instrument used in the experiments is from Ocean Optics Inc. (Dunedin, Florida, USA). The spectroscopic system features and experimental setup are similar to those reported in our preceding studies [25, 32, 33]. The optical fiber was connected to a QE65000 spectrometer with a spectral sensitivity between 350.64 and 1131.24 nm. However, the wavelength region of 400-1100 nm was utilized in the entire analysis. Three wells of the 6-well cell culture were used, and triplicate spectra were acquired from each well to guarantee the consistency of the measurements and obtain the average spectrum for the entire analyses. The acquisition parameters for the spectrometer including integration time, scan to average, and boxcar width was 17 ms, 8, and 3, respectively. Spectra Suite Software (Ocean Optics) was used to acquire and analyze the recorded spectra.

2.3. Data Analysis

MLR was used to estimate the number of cells for L929, DU145, and HeLa cell lines. MLR analysis is a multivariate statistical technique utilized to model biological processes. The MLR explores the relationship between two or more independent variables and one dependent variable [34]. The MLR model ensures the selective extraction of specific information from highly overlapping NIR spectra. The use of an MLR model with fewer wavelengths is usually more robust since increasing the number of selected wavelengths may increase the specificity of the model to a particular dataset. Collinear data are inappropriate for MLR calibrations. The selection of the wavelengths depends on the coefficient of determination, standard error of the cross-validation method, and a significance level (α) [35]. Before performing the MLR, all X and Y variables were weighted to 1, and α of 0.05 was used. In addition, the selection of wavelengths is depending on the significance of each wavelength in the MLR model which is considered ≤ 0.05. Calibration and full cross-validation methods were used to develop and validate the performance of the MLR models, respectively. For the model validation, the leverage correction method was performed which is strictly similar to the full cross-validation method in the MLR [36]. Statistical characteristics that are used to show the performance of the constructed MLR models were the standard error of estimation (SEE), coefficient of determination (R^2), and a significance level (α) [37]. The SEE is defined as the root mean square error of estimation (RMSEE) corrected for bias; RMSEE is a measure of the distribution of the validation samples around the regression line. The value of $\alpha < 5\%$ in the MLR model indicates the model elucidates more of the deviations of the response variable compared to random phenomena. This means that the model is significant, and the smaller α, the higher significance of the model is. MLR modelling was performed on raw

spectral data without using any pre-processing using the Unscrambler Software (version 10.4, Camo Software AS, Oslo, Norway).

PLSR was used as another method in this study for quantification of the cells using the NIR region 680-1050 nm. To optimize the PLSR modelling, various spectral pre-processing techniques were examined such as standard normal variate (SNV) [38], multiplicative scatter correction (MSC) [39], smoothing using Savitzky-Golay [40] with 2nd order polynomial filter with 21- and 11-points window, 1st and 2nd Savitzky-Golay derivative transformation with 15 points, and orthogonal signal correction (OSC) [41]. It should be mentioned that before model development, the spectra were mean-centered. The validation was performed internally using the leave-one-out cross-validation method. As quality determination for respective models, the coefficient of determination (R^2) and standard error of cross-validation (SECV) were utilized. The optimal number of factors or latent variables (LVs) in the model was decided based on the local minimum of the SECV. The PLSR modelling was performed using a commercially available multivariate analysis software Pirouette v4.5 (Infometrix Inc., USA).

3. Results and Discussion

Figure 1 shows the example of the raw spectra (400-1050 nm) of one cell line which is DU145 cultured in MEM medium with and without phenol red dye. The spectra were labelled as DU145 PR1-10/DU145 XPR1-10 which means the DU145 was cultured in the medium with/without phenol red (PR/XPR) with the number of cells ranging from 50,000 to 275,000. The arrow in Figure 1 denotes the direction of the spectra with increasing the number of cells. Each spectrum is the average of nine measurements of spectra from three wells of the 6-well cell culture plate. The measurement of the spectrum from each well was repeated three times in different locations to ensure the repeatability of the measurement and eliminate the uncertainties due to

sampling. It is imperative to mention that the produced absorbance in Figure 1 resulted from both absorbance and scattering processes of the light and it depends on the physical and chemical properties of the samples [42]. Therefore, the observed absorbance can occur when light is scattered by a specific number of cells. As can be seen from Figure 1, there are two peaks in the visible and NIR regions. The predominant peak in the visible region (558 nm) denotes the basic form of phenol red [43, 44]. Given that the culture medium is composed of water, a distinctive absorbance peak is observed at NIR wavelength (≈ 965-985 nm) is referred to as the second overtone of the water absorbance peak [45]. It is evident that the absorbance of the samples in the NIR region is less affected or predisposed to the color of the medium.

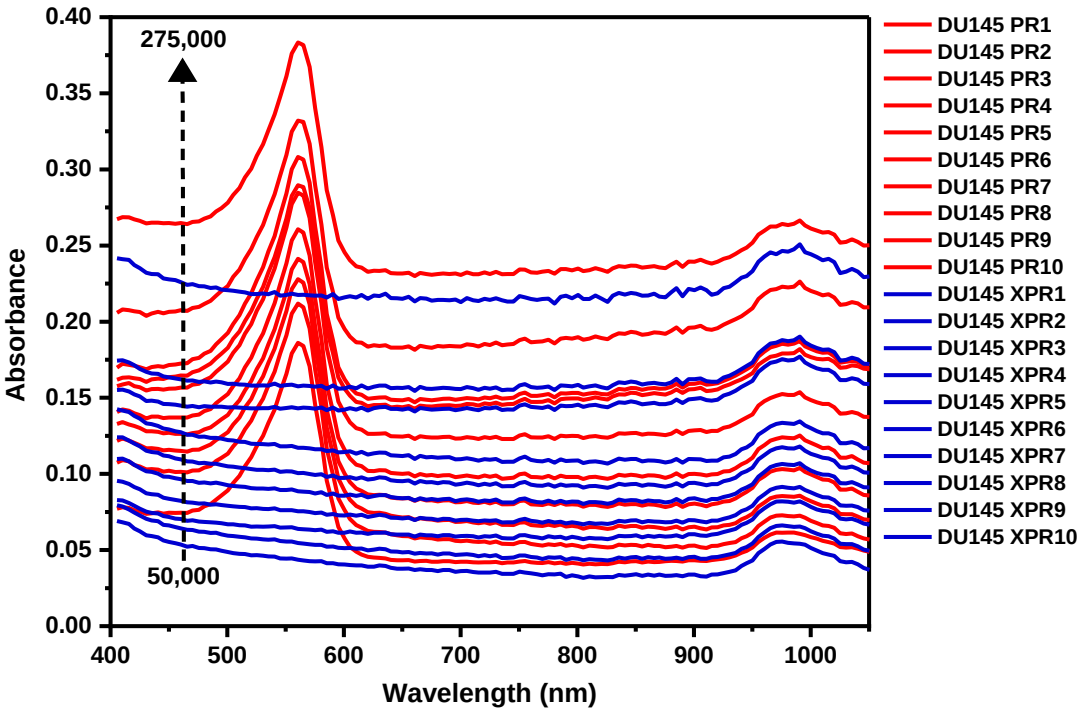


Figure 1. Vis-NIR absorbance spectra at region 400-1050 nm for DU145 cell line cultured in medium with phenol red (red color) and without phenol red (blue color) at the number of cells from 50,000 to 275,000.

NIR spectroscopy has the benefit of ease and expediency when applied to aqueous samples [46]. The use of NIR spectroscopy for cell cultivation has been recognized for the prediction of total or viable cell density and concentration of main metabolites. The peak absorbance values of the spectra at region 700-1050 nm for the three cell lines (L929, DU145, and HeLa) for 50,000 to 275,000 cells are presented in Supplementary Figure S1. In this figure, the bars represented the standard error of the mean (SE) of the peak intensity in the NIR region for each cell line at each number of cells. Each data point (mean $\pm$ SE) was calculated from nine recorded spectra that were obtained from three wells. Generally, it can be inferred that the peak absorbance values are consistent for the three cell lines. Nonetheless, there is a variation between the absorbance peak of the HeLa cell line grown in the presence and absence of phenol red in the culture medium at a high number of cells (250,000 and 275,000). This can be attributed to the relatively higher light scattering of HeLa cells compared to L929 and DU145 cells [25] and possibly due to a reduction in the emitted light at the detector. Light scattering minimizes the signal-to-noise ratio, thus intensifying the root mean square noise [47]. This may decrease the measurement precision of the absorbance spectra of the HeLa cell line. Furthermore, the variation in the peak absorbance in some number of cells may be because of errors by a person inexperienced with cell counting using the hemocytometer technique.

As reported in our previous work [33], the results showed that culturing different types of cells in medium with phenol red resulted in diminished interclass distance, and decreased the potential for discrimination between the three cell types (in PCA analysis). On the other hand, this has a benefit of easier quantification by creating general regression model that can be used for quantification of cells despite of their type. The aim of investigating the cells cultured in a medium with and without phenol red is to show the implication of the phenol red in the culture medium on

cells in the NIR region. The NIR absorbance spectra of the cells are less affected or predisposed to the color of the medium due to phenol red. Furthermore, it was showed that both water content and molecular structure of the cell culture (media and cells) are affected by staining.

In general, the pre-processing of the spectra is extensively used to eliminate/minimize the unwanted artefacts from the spectra. For PLSR/MLR models, various pre-processing methods and their combinations are frequently examined to find the optimum performance in terms of the minimum prediction error. In some cases, the scattering data might offer further merit associated with the physical composition of materials which could augment the behavior of the model [48]. As proved from the MLR and PLSR models in this study using NIR spectra at region 700-1050 nm of the three cell lines, the scattering may comprise vital details related to the feature of interest. Therefore, pre-processing techniques that diminish the scattering data (such as MSC and SNV) or their combination may result in badly processing models.

This study is aimed to build an algorithm model for quantification of the number of cells of L929, DU145, and HeLa cell lines. The MLR is appropriate for systems with a few variables [49] and is preferred on account of its non-complexity and easy operation [50]. The analysis was performed by applying MLR using the absorbance datasets of L929, DU145, and HeLa cell lines. The spectral data used was restricted to the 700-1050 nm wavelength range as the relative absorbance measurements outside this range were related to the color of the medium due to the presence of phenol red or too noisy. Before building the MLR models, the selected wavelengths (SWs) were obtained by the regression coefficients of PLSR models, and loadings plots of the PCA analysis. These wavelengths are considered to provide more contribution to the calibration of the MLR model [51].

In the case of including phenol red in the culture medium, the wavelengths that were selected for quantifying the number of the cells in the NIR region were 775, 841, 955 nm for L929 and for both DU145 and HeLa were 700, 801, and 976 nm. Individual (or local) MLR models were constructed using these wavelengths for the three cell lines. Besides, the local MLR models in the case of growing the cells in the media without phenol red were developed. The SWs were 736, 941, and 1015 nm for L929; 736, 775, 805, and 906 nm for DU145; 810, 861, 960, and 986 nm for the HeLa cell line. Table 1 summarizes the local MLR models and the statistics of the full cross validation (sometimes called Leave-One-Out "LOO" cross validation) [52] in the presence and absence of the phenol red. In addition, the predicted versus the actual number of cells using the calibration and cross-validation methods obtained from the local MLR models are shown in Figure 2(a)-(f) for L929, DU145, and HeLa cell lines. In this figure, the R^2 and standard errors of calibration and cross-validation (R_C^2, R_{CV}^2, SEC, and SEE) for each model are included.

Table 1. MLR local models and cross-validation statistics for each L929, DU145, and HeLa cell line to predict the number of cells in case of presence/absence of phenol red in culture media.

With phenol red											
L929				**DU145**				**HeLa**			
SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α
775, 841, 955	0.99	5,995	2.22E-08	700, 801, 976	0.99	7,498	7.01E-08	700, 801, 976	0.98	10,657	5.86E-07
Without phenol red											
L929				**DU145**				**HeLa**			
SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α	SWs (nm)	R_{CV}^2	SEE (Cells)	α
736, 941, 1015	1	4,771	1.29E-08	736, 775, 805, 906	1	3,363	2.33E-08	810, 861, 960, 986	0.95	17,442	9.44E-05

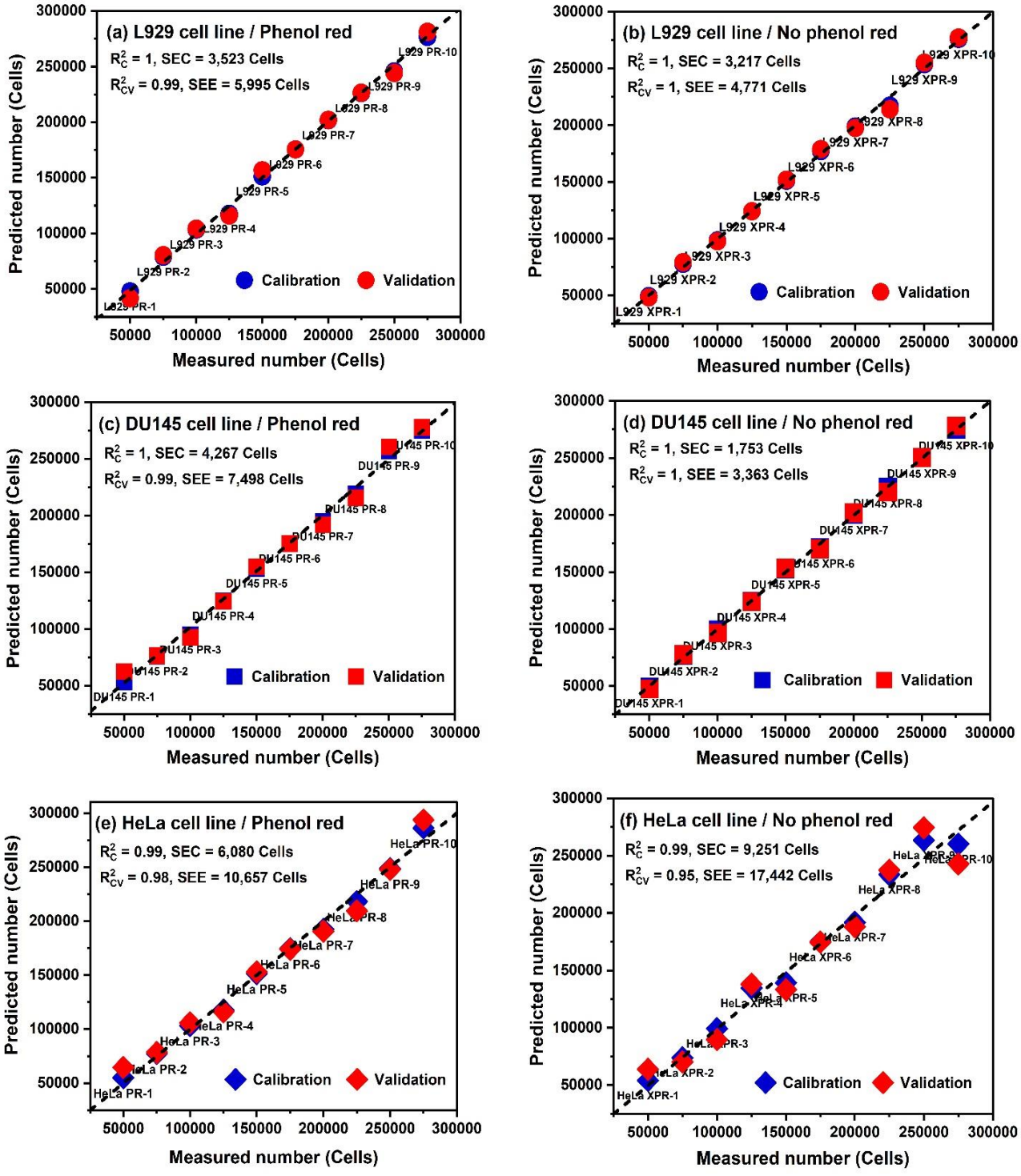


Figure 2. Predicted versus measured values of calibration (red) and cross-validation (blue) results of local MLR models for quantification the cells: (a) and (b) for L929, (c) and (d) for DU145, and (e) and (f) for HeLa cell line in case of presence/absence phenol red in culture media, respectively.

A general (or global) MLR model for estimating the number of cells of the three cell lines together (L929, DU145, and HeLa) in the presence (using SWs: 736, 775, and 805 nm) and absence (using SWs: 745, 810, 861, 900, 960, 986, and 1010 nm) of the phenol red in the culture media was constructed. The samples were designated as "PR" and "XPR" which means the presence and absence of phenol red, respectively. Table 2 shows the global MLR models and the

Table 2. MLR global models and cross-validation statistics for L929, DU145, and HeLa cell lines to predict the number of cells in case of presence/absence of phenol red in culture media.

With phenol red			
SWs (nm)	R^2_{CV}	SEE (Cells)	α
736, 775, 805	0.96	14,661	3.47E-20
Without phenol red			
SWs (nm)	R^2_{CV}	SEE (Cells)	α
745, 810, 861, 900, 960, 986, 1010	0.93	19,138	1.22E-14

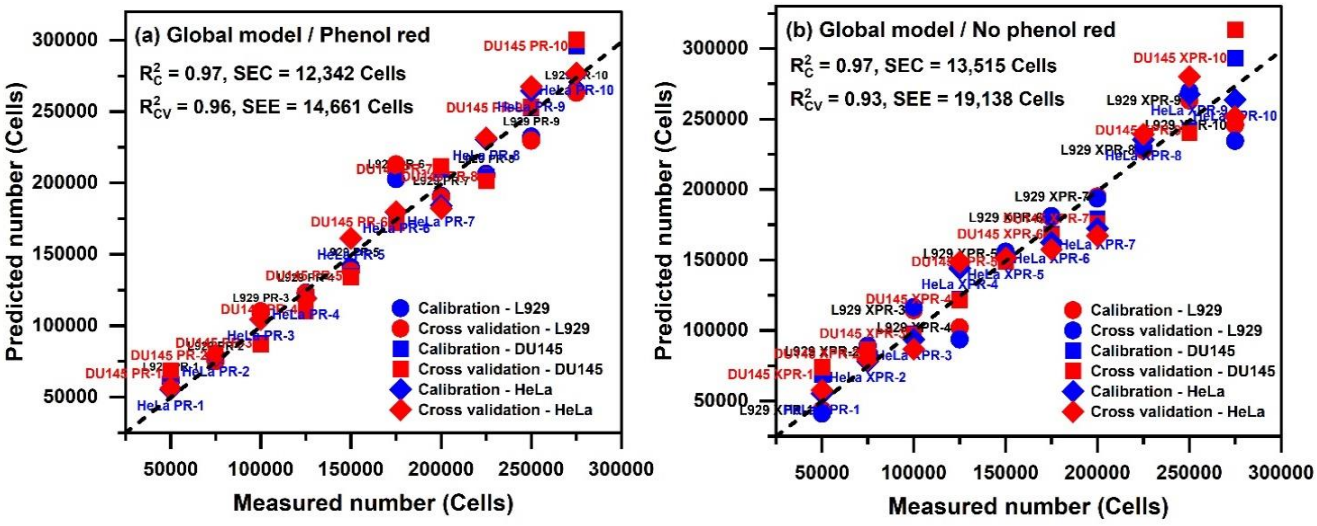


Figure 3. Predicted versus measured values of calibration (red) and cross-validation (blue) results of global MLR models for quantification of the L929, DU145, and HeLa cells grown in the medium: (a) with phenol red and (b) without phenol red.

statistics of the cross-validation method in the presence and absence of the phenol red. Figure 3 displays the global MLR models for predicting the number of cells. The obtained values of R^2_{CV} and SEE of the MLR models suggested its good predictive capacity and the appropriateness of the selected wavelengths in the spectral region 700-1050 nm without any spectral pre-treatments.

The PLSR modelling was performed in two ways: firstly, a local PLSR model was created for each cell line separately (L929, DU145, and HeLa) that is best adapted for a specific cell line. Secondly, a global PLSR model was developed for quantification of the number of cells irrespective of the cell line type. In both cases, models were built separately for the cells grown in media with and without phenol red to discern the effects of phenol red on cell culture and quantification. Quantitative modelling was first performed using PLSR analysis on datasets of spectra for each cell line separately in order to build local models assuming that the accuracy of prediction would be the highest and allow assessment of effects of phenol red treatment. Table 3 presented the PLSR model and validation statistics of testing different pre-processing techniques for local PLSR models in both cases with and without the presence of phenol red treatments. From Table 3, it can be observed that the removal of additive and multiplicative effects (scatter correction techniques: SNV, MSC, and $1^{st}/2^{nd}$ derivatives) led to poorer model performance. This means that the baseline variations contain significant information for the quantification of cells, which is in agreement with previous research reports about the quantifications of fat (globules), protein (micelles), and somatic cells [53, 54]. Based on the quality measures for respective PLSR models, it can be established that the best accuracy was obtained with the OSC pre-processing method using one component, irrespective of phenol red treatment and cell line type as shown in Table 3. The validation results showed R^2 higher than 0.99 using the OSC pre-processed spectra. The treatment of phenol red improved the accuracy of cells number prediction in the case of the

cancer cells (DU145: SECV = 250 cells and HeLa: SECV = 271 cells). However, for L929 normal cells, the accuracy of the PLSR models was enhanced in the case of not including the phenol red in the culture medium (SECV = 24 cells). In addition, the accuracy of quantification was poorest for HeLa cells, and the best in the case of L929 cells. Figure 4 shows the results of calibration and leave-one-out cross-validation methods of the local PLSR models developed using the pre-processing by OSC (one component) for mean-centered spectra. This figure shows a good agreement between the measured and predicted number of cells by PLSR model. Figure 5

Table 3. Local PLSR models and cross-validation statistics for each L929, DU145, and HeLa cell line to predict the number of cells in case of presence/absence phenol red in culture media.

Case	**Pre-processing method**	**L929**			**DU145**			**HeLa**		
		LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)
With phenol red	None	5	0.99	12,059	3	0.99	11,685	4	0.99	8,064
	Smoothing (11 points)	2	0.99	12,517	4	0.99	9,348	3	0.99	11,275
	SNV	4	0.89	34,022	6	0.86	37,709	5	0.79	44,570
	MSC	5	0.58	73,835	2	0.83	41,045	5	0.8	43,929
	1st Derivative (15 points)	7	0.69	55,121	6	0.93	27,399	6	0.89	34,213
	2nd Derivative (15 points)	3	0.76	46,852	3	0.96	21,256	2	0.78	45,231
	OSC	**2**	**1**	**137**	**1**	**1**	**250**	**1**	**1**	**271**
Case	**Pre-processing method**	**L929**			**DU145**			**HeLa**		
		LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)	LVs	R^2_{CV}	SECV (Cells)
Without phenol red	None	2	0.99	5,570	5	0.96	23,077	2	0.96	21,032
	Smoothing (21 points)	2	0.99	5,625	4	0.97	19,351	2	0.96	20,993
	SNV	2	0.78	45,562	3	0.96	18,939	1	0.78	46,253
	MSC	2	0.78	45,525	3	0.96	19,008	1	0.76	47,500
	1st Derivative (15 points)	2	0.62	60,557	3	0.97	18,751	1	0.78	44,847
	2nd Derivative (15 points)	1	0.68	54,716	1	0.92	27,577	3	0.74	48,440
	OSC	**3**	**1**	**24**	**1**	**1**	**1,364**	**1**	**0.99**	**7,444**

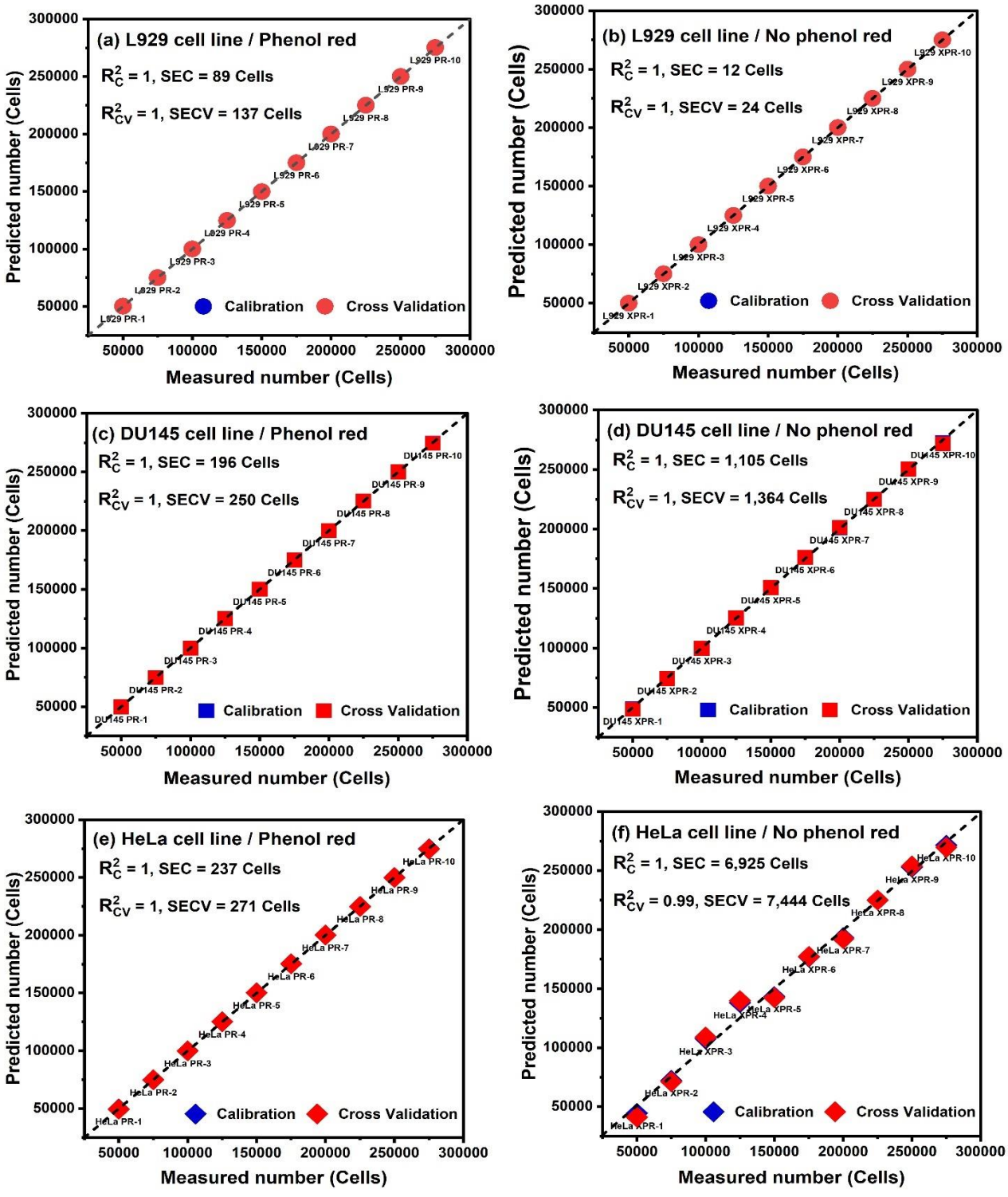


Figure 4. Predicted versus measured values of calibration (red) and cross-validation (blue) results of local PLSR models for quantification of cells: (a) and (b) for L929, (c) and (d) for DU145, and (e) and (f) for HeLa cell line in case of presence/absence phenol red in culture media, respectively.

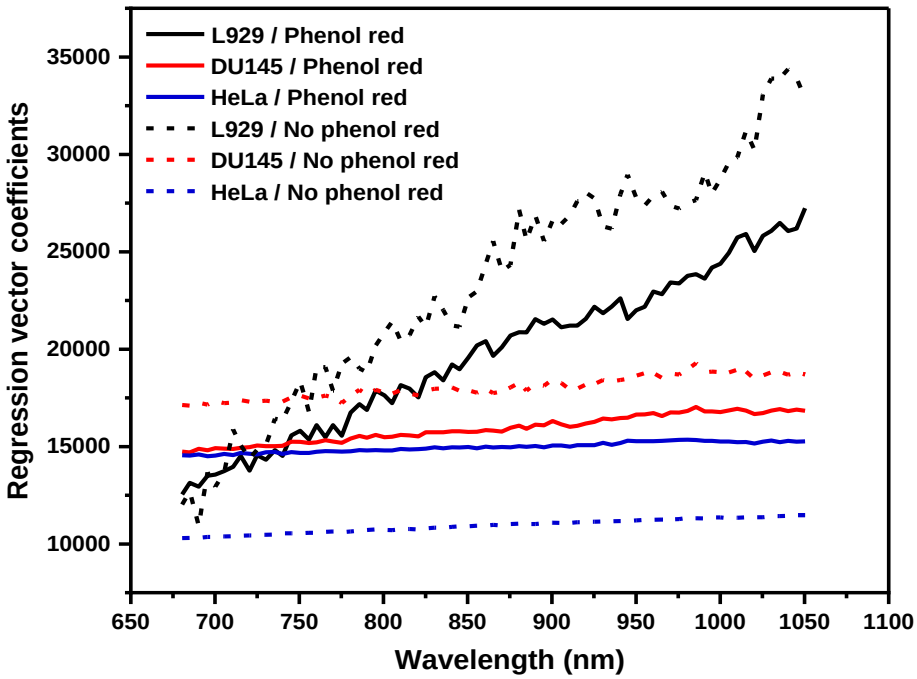


Figure 5. Regression vector coefficients of local PLSR models obtained from the best accuracy of prediction of the number of cells of L929, DU145, and HeLa cell lines in case of presence/absence of phenol red in culture media.

shows the regression vector coefficients of the PLSR models developed from Figure 4 for each cell line. The lack of any marked peaks validates the fact that the scatter plays the important role in the PLSR modelling instead of the absorbance [52]. It can be observed from Figure 5 that the cells' quantification is related to the variations of the baseline, in other words, the quantification of the cells is based on the effects of light scattering by the cells. In addition, from Figure 5, in the case of both DU145 and HeLa cell lines (cancer cells), the baseline effects are mostly additive resulting in the shift of the baseline of the spectra. However, in the case of the L929 cell line (normal cells), there are also multiplicative effects resulting in the change of the slope of the baseline. Interestingly, the use of phenol red influences the scattering of light by cells which is attested by the shifting in the regression vectors, but the shape is quite preserved. It can be deduced that the building of the local PLSR models for each cell line resulted in high accuracy of the

quantification of the cells. The quantification of the cells was better when L929 cells were grown in the medium without phenol red, while on the opposite, the quantification accuracy for DU145 and HeLa extremely improved when the phenol red was included in the culture media. These findings demonstrate the influence of phenol red on light-scattering properties, which are different and dependent on the cell type.

Two global PLSR models in the case of the presence and absence of phenol red were developed that allowed quantification of cells irrespective of the cell type. Various pre-processing techniques were also used in order to extract the information related to the number of cells and enhance the performance of the global model. The results of the cross-validation method of the developed global PLSR models using different pre-processing methods are summarized in

Table 4. Global PLSR models and cross-validation statistics for L929, DU145, and HeLa cell lines to predict the number of cells in case of presence/absence of phenol red in culture media.

Case	Pre-processing method	LVs	R^2_{CV}	SECV (Cells)
With phenol red	None	5	0.98	13,007
	Smoothing	7	0.98	13,694
	SNV	3	0.73	50,343
	MSC	3	0.62	60,740
	1st Derivative	5	0.88	34,176
	2nd Derivative	7	0.75	49,802
	OSC	**1**	**0.99**	**7,754**
Case	**Pre-processing method**	**LVs**	R^2_{CV}	**SECV (Cells)**
Without phenol red	None	5	0.93	26,577
	Smoothing	5	0.94	25,447
	SNV	3	0.79	43,826
	MSC	3	0.8	43,606
	1st Derivative	5	0.88	34,380
	2nd Derivative	2	0.82	41,279
	OSC	**1**	**0.97**	**17,008**

Table 4. It is clear from Table 4 that also for both global models, OSC pre-processing using one component produced the best accuracy of quantification and resulted in a simple model with only one latent variable. Compared to the established local models for each cell line, the global models showed a significant decrease in the accuracy of predicting the number of cells. This implies that local modelling considers the specific characteristics of each cell line which is a better choice in quantitative modelling. Furthermore, from Table 4, the phenol red in the cultured medium resulted in improved accuracy of cells' quantification. The relationship between the measured and predicted number of cells using the global PLSR models developed using the pre-processed spectra by the OSC method for both cases with and without phenol red treatments is shown in Figures 6(a) and (b), respectively. It can be inferred that the phenol red has effects on the cells, in the terms of diminishing the differences between the individual cell lines. This is in good agreement with the findings of our preceding study, wherein in the case of discrimination between the three cell lines, cells cultured in media including phenol red showed diminished class distances in discriminative models [33]. As can be seen from Figure 6(a), it is evident that phenol red in the cell culture media led to a diminishment of the error and decreased differences between the three cell lines. This confirms the influence of the phenol red on the characteristics of the cells and their subsequent interaction with near-infrared light. On the other hand, for the cells grown in the media without phenol red as presented in Figure 6(b), there is a small error between the measured and predicted number of cells at a small number of cells. Whereas, at a higher number of cells, the error increases and the distinction between the cell lines is more pronounced. The regression vector coefficients of the PLSR global models of the three cell lines (L929, DU145, and HeLa) are presented in Figure 7. Apparently, the regression vectors of the global PLSR model established for the case of cells grown in media with phenol red (solid line) show a higher magnitude of coefficients. This can be

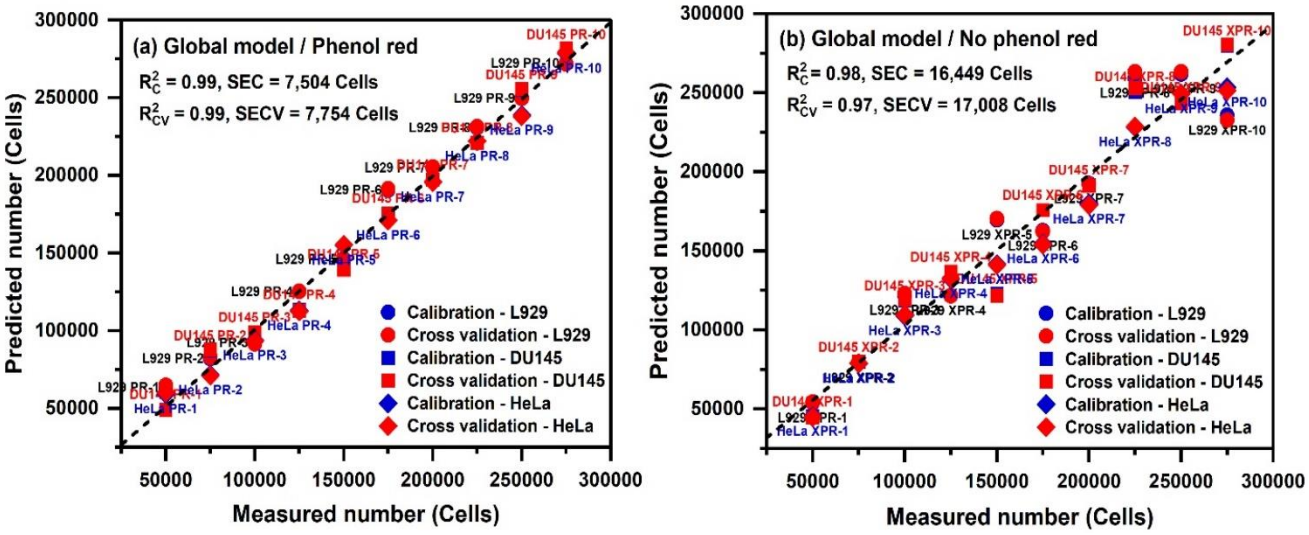


Figure 6. Predicted versus measured values of calibration (red) and cross-validation (blue) results of global PLSR models for quantification of the L929, DU145, and HeLa cells grown in the medium: (a) with phenol red and (b) without phenol red.

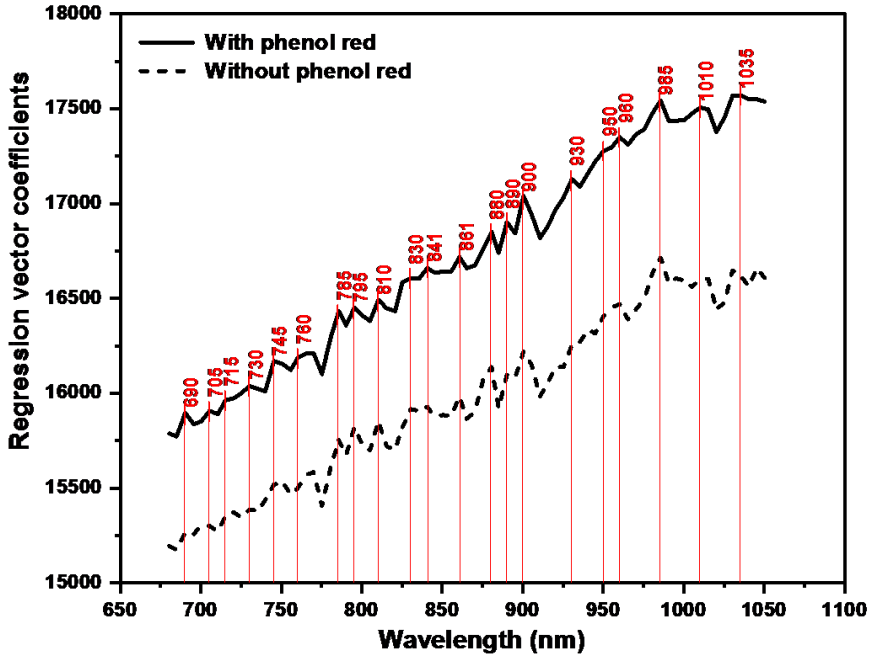


Figure 7. Regression vector coefficients s of global PLSR models for quantification of L929, DU145, and HeLa cells grown in the media with and without phenol red.

attributed to the increased scattering of light by the cells. However, the shape of the regression coefficients is very similar for both models. This revealed that both global models are based on

variations of the spectral baseline in connection to the number of cells. Despite the dominant effect of scattering, absorbance bands can be observed in the regression coefficients and almost comparable absorbance bands characterize the regression coefficients of the two global models.

In the NIR spectral region, particularly at the region below 1000 nm, the scattering strongly dominates over absorption. It was reported in various works that this region can be used as a source for quantitative information on light-scattering of colloidal particles [52, 55-57] which can explain the findings acquired in this work. The main factors affecting the scattering of a single particle or a collection of particles are the refractive index difference between the particles and the matrix, the size of the particle, and the wavelength of the incident light [58]. It is most likely that the refractive index of the matrix (the medium in which the cells were grown, and the highest component of the medium is water) is affected by the presence of phenol red.

For more investigation, we examined the difference in influential variables between the created local PLSR models depending on the presence (PR) and absence (XPR) of phenol red in the media where the cells were grown. For this purpose, the regression vector coefficients of the corresponding models were detrended to eliminate the slope, and then the difference between the regression vector coefficients of PR and XPR models was calculated. The plots of regression vector coefficients after detrending and the result of their subtraction are presented in Figures 8-10 for L929, DU145, and HeLa PLSR models, respectively. Interestingly, the highest differences in regression coefficients in all models are located in the same region which is known as the 2nd overtone of water (approximately around 900-1000 nm). The largest differences in this region can be observed at bands: 915 and 945 nm for L929 cells, 955 and 986 nm for DU145 cells, and 930, 945, 970, and 976 nm for HeLa cells. Obviously, there is a difference in these particular bands in the terms of their importance as influential variables for building the PLSR model. This difference

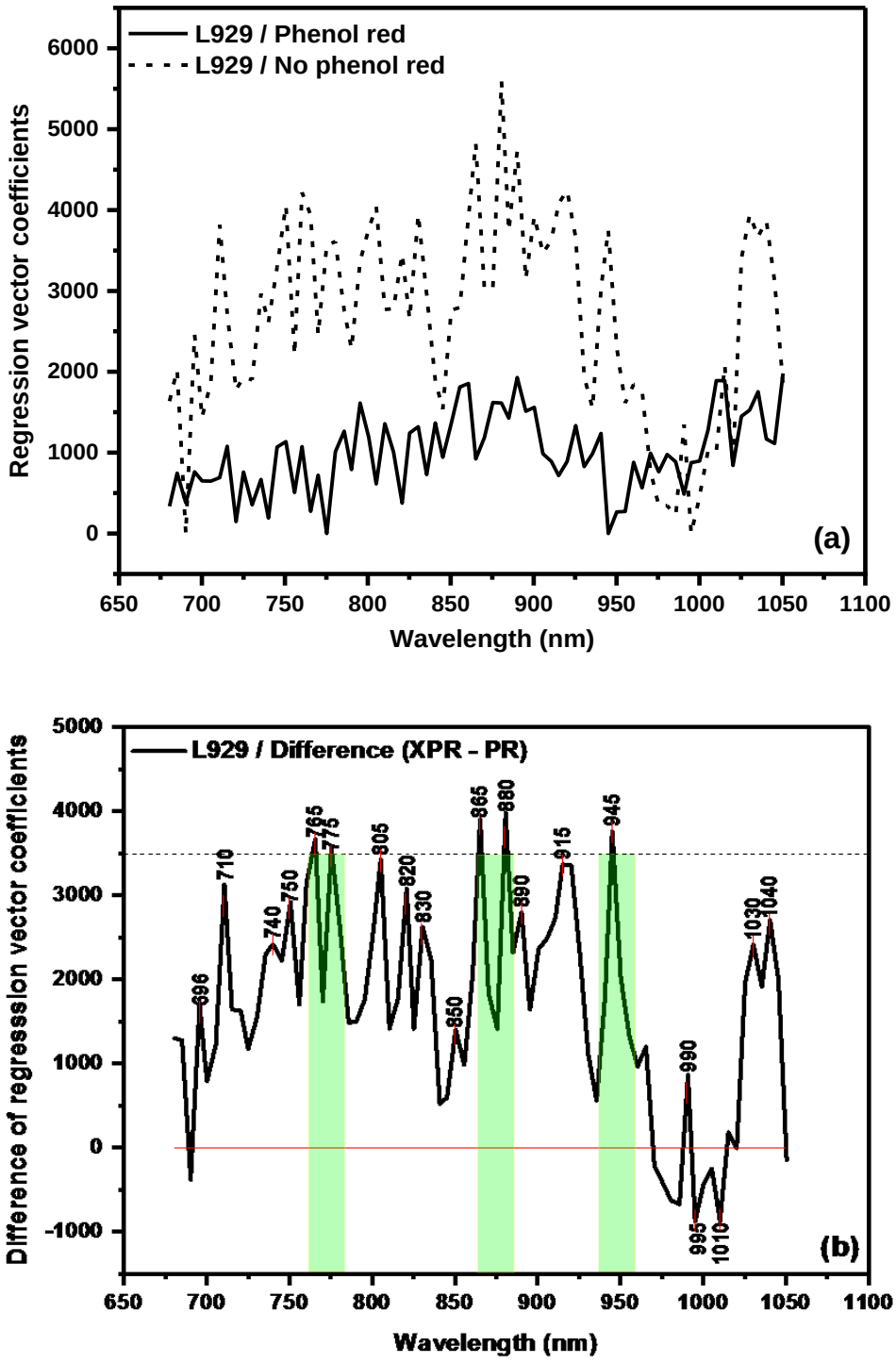


Figure 8. Detrended regression vector coefficients of PLSR models for L929 cells quantification built using spectral data from cells in media with phenol red (PR) and without phenol red (XPR) (a) and their difference (b).

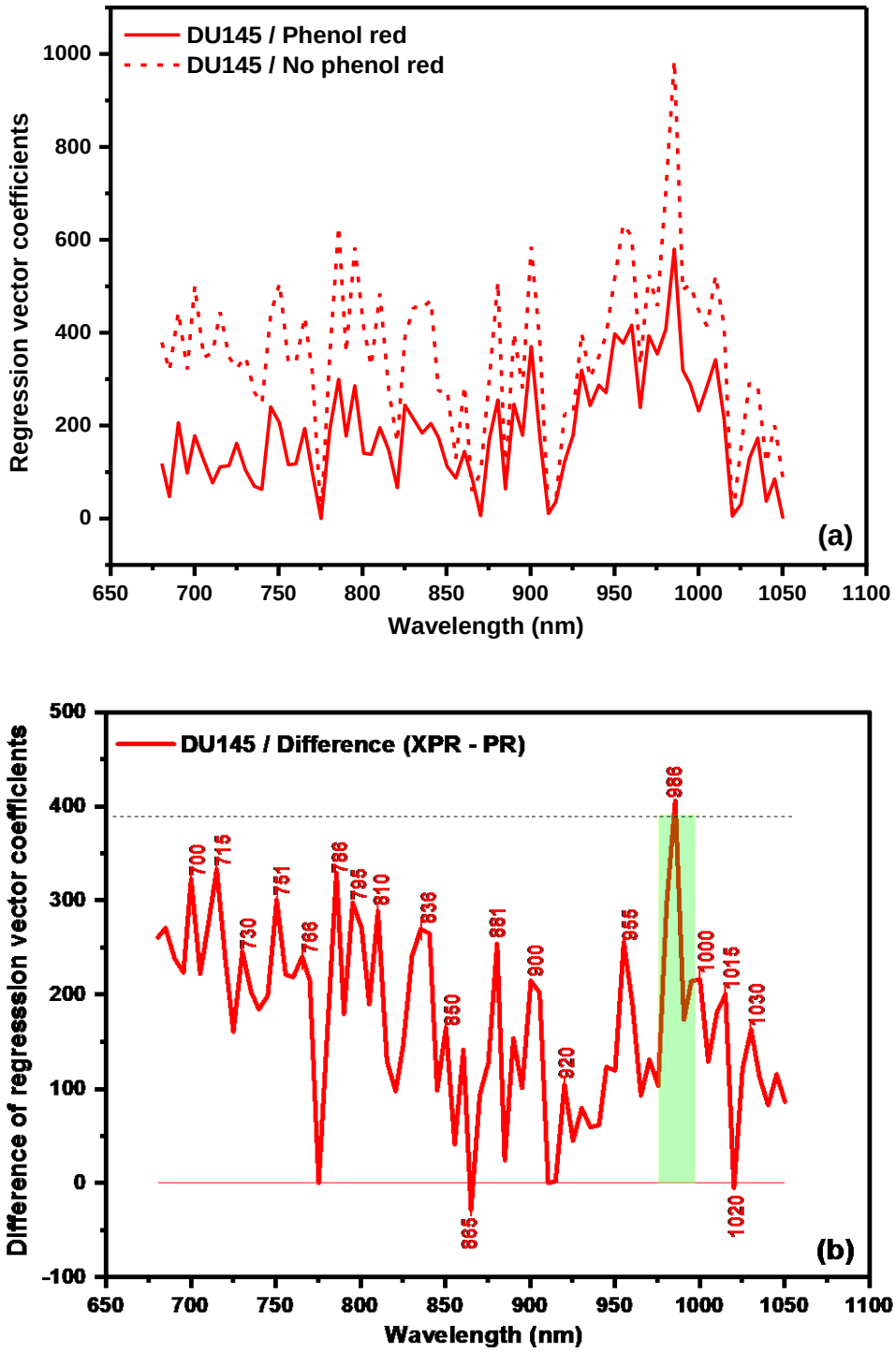


Figure 9. Detrended regression vector coefficients of PLSR models for DU145 cells quantification built using spectral data from cells in media with phenol red (PR) and without phenol red (XPR) (a) and their difference (b).

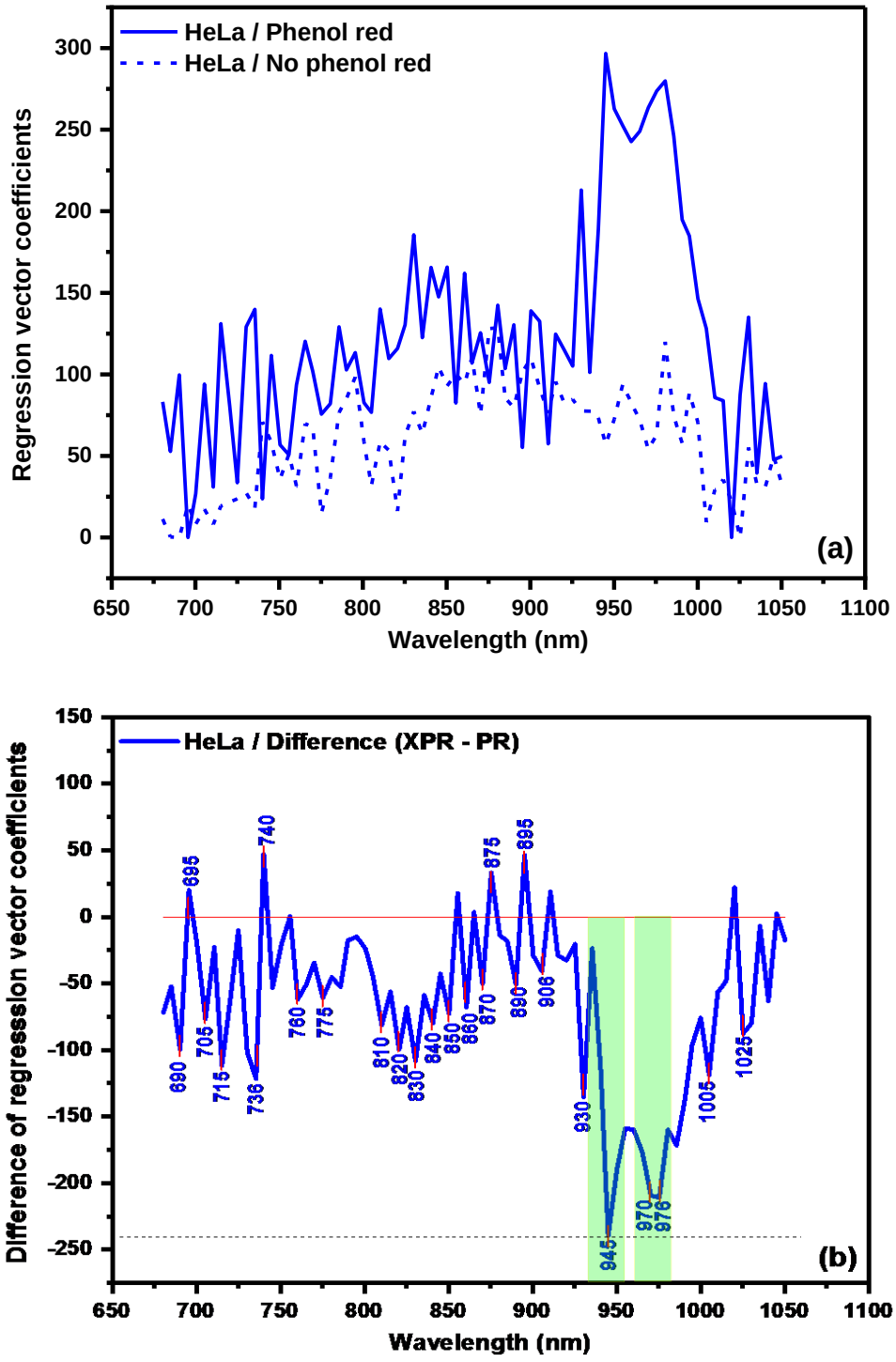


Figure 10. Detrended regression vector coefficients of PLSR models for HeLa cells quantification built using spectral data from cells in media with phenol red (PR) and without phenol red (XPR) (a) and their difference (b).

depends on whether the cells were grown in the absence or presence of phenol red and is specific for each cell type. The assignments of the bands that are of importance (the highest bands) for the accuracy of the PLSR models depending on the presence of phenol red in the culture media are given in Table 5. Since the bands' positions were in the 2nd overtone region of water, their assignments were found based on the calculated position in the 1st overtone of water as presented in Table 5. The bands shown in Table 5 are the 2nd overtones of the well-known Water Matrix Coordinates "WAMACs" in aquaphotomics (water absorbance bands attributable to vibrations of the particular water molecular species) [59]. This confirms that the presence of phenol red has an influence on the cells' culture media by changing the water structure. Further, considering that different bands are affected by the type of the cells, the water structure of the media is a result of both influences of cell metabolites and phenol red. As a result, for the DU145 cells, the specific bands 955 and 986 nm, i.e., water dimers and tetramers are affected; in the case of HeLa cells are water trimers, solvation shells, free and quasi-free water molecules, while in the case of L929

Table 5. Assignments of the bands found to be of importance for cells quantification accuracy using PLSR analysis, depending on the presence of phenol red in the cell media.

Band position in 2nd overtone of water (nm)	**Calculated position in 1st overtone of water (1300 - 1600 nm)**	**Assignment / WAMACs [59-61]**
915	1372.5	C3 – $\nu_1 + \nu_3$: Water symmetrical stretching vibration and water asymmetric stretching vibration
930	1395	C5 – Water molecules confined in the local field of ions
945	1417.5	C5 – Free water molecules, S_0
955	1432.5	C6 – S1: Water molecules with 1 hydrogen bond
970	1455	C8 – Water solvation shell, $OH^-(H_2O)_{4,5}$
976	1464	C9 – Water molecules with 2 hydrogen bonds, S_2
986	1479	C10 – Water molecules with 3 hydrogen bonds, S_3

cells, in particular free water molecules. It is important to point out that in the case of L929 cells, there is a small difference between the regression vector coefficients at hydrogen-bonded water wavelengths ($>$ 945 nm), and this makes a stark difference when compared to DU145 and HeLa cells where the largest differences were observed exactly in hydrogen-bonded water. Since in all cases the phenol red caused changes in the molecular structure of water, the refractive index must also change. This can be explained by the refractive index dependence on the molecular structure [62] of the matrix, resulting in different scattering properties of each cell line media and different quantification outcomes.

The results of this study should also bring more realization to the so-called "Observer Effect", an important and frequent experience in biomedical research, where the tools used to facilitate observation may exert unrecognized and often undesired effects on the underlying biology [63]. While the phenol red is routinely used in laboratory practice, its effects on the system the researchers are working with are still unrecognized. This study showed the effective treatment of the phenol red dye towards better quantification of the cultured cells. Furthermore, the molecular structure of water is not yet recognized as an important experimental variable and parameter that should be considered in research studies. It was shown here that the molecular structure of water was altered by the presence of phenol red in the cells' culture medium. The findings of this study, therefore, enhance awareness of these issues and revealed the informative power of NIR spectroscopy and aquaphotomics in this regard.

4. Conclusion

This work presented the application of NIR absorbance spectroscopy data to quantify the number of cultured cells. Three cell lines including cancerous cells (DU145 and HeLa) and normal cells (L929) were used. The number of cells for each cell line was ranging from 50,000 to 275,000

with an interval of 25,000. The Vis-NIR absorbance spectra (400-1100 nm) were measured for each number of cells cultured in media with and without phenol red. To quantify the number of cells for the three cell lines, the local and global regression models using multiple linear regression (MLR) and partial least squares regression (PLSR) were developed in the NIR region (680-1050 nm). It was revealed that the MLR and PLSR models for the quantification of the cells produced high prediction accuracy with $R^2 \geq 93\%$. It was noticed that the accuracy of the local and global PLSR models, especially for cancer cells was higher in the case of cultured cells in media including phenol red dye. This can be an important finding in this work which showed the implication of phenol red in cell culture media towards the quantification of the number of cancerous cells. The proposed method in this study can be considered a novel and rapid approach for quantifying the number of cancerous cells.

Acknowledgment

The authors with gratitude acknowledge the financial support of this research by the Ministry of Higher Education Malaysia–Fundamental Research Grant Scheme (Grant No. FRGS/1/2020/STG07/USM/02/8).

References

1. Sorvina, A., et al., *Lipid profiles of prostate cancer cells.* Oncotarget, 2018. **9**(85): p. 35541.

2. Loud, J.T. and J. Murphy. *Cancer screening and early detection in the 21st century*. in *Seminars in oncology nursing*. 2017. Elsevier.

3. Vembadi, A., A. Menachery, and M.A. Qasaimeh, *Cell cytometry: Review and perspective on biotechnological advances.* Frontiers in bioengineering and biotechnology, 2019. **7**: p. 147.

4. Friberg, S. and S. Mattson, *On the growth rates of human malignant tumors: implications for medical decision making.* Journal of surgical oncology, 1997. **65**(4): p. 284-297.

5. Frangioni, J.V., *New technologies for human cancer imaging.* Journal of clinical oncology, 2008. **26**(24): p. 4012.

6. Shern Khoo, A.B., et al., *Comparative Analyses of Tumour Volume Doubling Times for Periocular and Non-periocular Head and Neck Basal Cell Carcinomas.* Acta dermato-venereologica, 2019. **99**(13): p. 1266-1269.

7. O'Brien, J., H. Hayder, and C. Peng, *Automated quantification and analysis of cell counting procedures using ImageJ plugins.* Journal of visualized experiments: JoVE, 2016(117).

8. Ongena, K., et al., *Determining cell number during cell culture using the Scepter cell counter.* Journal of visualized experiments: JoVE, 2010(45).

9. Brown, J.Q., et al., *Advances in quantitative UV–visible spectroscopy for clinical and pre-clinical application in cancer.* Current opinion in biotechnology, 2009. **20**(1): p. 119-131.

10. Solano, R.P., et al., *An experimental and theoretical approach to the study of the photoacoustic signal produced by cancer cells.* AIP Advances, 2012. **2**(1): p. 011102.

11. Krediet, C.J., et al., *Rapid, precise, and accurate counts of symbiodinium cells using the guava flow cytometer, and a comparison to other methods.* PLoS One, 2015. **10**(8): p. e0135725.

12. Blasi, T., et al., *Label-free cell cycle analysis for high-throughput imaging flow cytometry.* Nature communications, 2016. **7**(1): p. 1-9.

13. Grishagin, I.V., *Automatic cell counting with ImageJ.* Analytical biochemistry, 2015. **473**: p. 63-65.

14. Mölder, A., et al., *Non- invasive, label- free cell counting and quantitative analysis of adherent cells using digital holography.* Journal of microscopy, 2008. **232**(2): p. 240-247.

15. Talebian, S. and M. Javanmard, *Compact and automated particle counting platform using smartphone-microscopy.* Talanta, 2021. **228**: p. 122244.

16. Sun, T. and H. Morgan, *Single-cell microfluidic impedance cytometry: a review.* Microfluidics and Nanofluidics, 2010. **8**(4): p. 423-443.

17. von Bartheld, C.S., J. Bahney, and S. Herculano- Houzel, *The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting.* Journal of Comparative Neurology, 2016. **524**(18): p. 3865-3895.

18. Trunfio, N., et al., *Characterization of mammalian cell culture raw materials by combining spectroscopy and chemometrics.* Biotechnology progress, 2017. **33**(4): p. 1127-1138.

19. Oshima, Y., et al., *Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy.* Journal of biomedical optics, 2010. **15**(1): p. 017009.

20. Dai, W.Y., S. Lee, and Y.C. Hsu, *Discrimination between oral cancer and healthy cells based on the adenine signature detected by using Raman spectroscopy.* Journal of Raman Spectroscopy, 2018. **49**(2): p. 336-342.

21. Shin, H., et al., *Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes.* ACS nano, 2020. **14**(5): p. 5435-5444.

22. Chan, J.W., et al., *Nondestructive identification of individual leukemia cells by laser trapping Raman spectroscopy.* Analytical chemistry, 2008. **80**(6): p. 2180-2187.

23. Talari, A., et al., *Raman spectroscopic analysis differentiates between breast cancer cell lines.* Journal of Raman Spectroscopy, 2015. **46**(5): p. 421-427.

24. Del Mistro, G., et al., *Surface-enhanced Raman spectroscopy of urine for prostate cancer detection: a preliminary study.* Analytical and Bioanalytical Chemistry, 2015. **407**(12): p. 3271-3275.

25. Abd Ghani, K., et al., *VIS–NIR spectral signature and quantitative analysis of HeLa and DU145 cell line.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019. **222**: p. 117241.

26. Sakudo, A., *Near-infrared spectroscopy for medical applications: Current status and future perspectives.* Clinica Chimica Acta, 2016. **455**: p. 181-188.

27. Geladi, P. and B.R. Kowalski, *Partial least-squares regression: a tutorial.* Analytica chimica acta, 1986. **185**: p. 1-17.

28. Martens, H. and M. Martens, *Multivariate analysis of quality: an introduction*. 2001: John Wiley & Sons.

29. Geladi, P. and E. Dåbakk, *Computational methods and chemometrics in near infrared spectroscopy.* 2017.

30. Shahmirani, S., E. Vasheghani Farahani, and J. Ghasemi, *Development of a model to predict partition coefficient of organic pollutants in cloud point extraction process.* Annali di Chimica: Journal of Analytical, Environmental and Cultural Heritage Chemistry, 2006. **96**(5- 6): p. 327-337.

31. Carrascal, L.M., I. Galván, and O. Gordo, *Partial least squares regression as an alternative to current regression methods used in ecology.* Oikos, 2009. **118**(5): p. 681-690.

32. Amran, E.N., et al., *Potential colorimetric detection of cancer cells using Phenol Red.* Photodiagnosis and photodynamic therapy, 2019. **27**: p. 380-384.

33. Raypah, M.E., et al., *Integration of near-infrared spectroscopy and aquaphotomics for discrimination of cultured cancerous cells using phenol red.* Chemometrics and Intelligent Laboratory Systems, 2022. **227**: p. 104611.

34. Jobson, J., *Multiple linear regression*, in *Applied multivariate data analysis*. 1991, Springer. p. 219-398.

35. Berntsson, O., G. Zackrisson, and G. Östling, *Determination of moisture in hard gelatin capsules using near-infrared spectroscopy: applications to at-line process control of pharmaceutics.* Journal of pharmaceutical and biomedical analysis, 1997. **15**(7): p. 895-900.

36. Esbensen, K.H., et al., *Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design*. 2002: Multivariate Data Analysis.

37. Farhadi, S., et al., *Modeling of paclitaxel biosynthesis elicitation in Corylus avellana cell culture using adaptive neuro-fuzzy inference system-genetic algorithm (ANFIS-GA) and multiple regression methods.* PloS one, 2020. **15**(8): p. e0237478.

38. Barnes, R., M.S. Dhanoa, and S.J. Lister, *Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.* Applied spectroscopy, 1989. **43**(5): p. 772-777.

39. Geladi, P., D. MacDougall, and H. Martens, *Linearization and scatter-correction for near-infrared reflectance spectra of meat.* Applied spectroscopy, 1985. **39**(3): p. 491-500.

40. Savitzky, A. and M.J. Golay, *Smoothing and differentiation of data by simplified least squares procedures.* Analytical chemistry, 1964. **36**(8): p. 1627-1639.

41. Wold, S., et al., *Orthogonal signal correction of near-infrared spectra.* Chemometrics and Intelligent laboratory systems, 1998. **44**(1-2): p. 175-185.

42. Blanco, M., et al., *Effect of data preprocessing methods in near-infrared diffuse reflectance spectroscopy for the determination of the active compound in a pharmaceutical preparation.* Applied Spectroscopy, 1997. **51**(2): p. 240-246.

43. Guo, Y., et al., *A pH-responsive colorimetric strategy for DNA detection by acetylcholinesterase catalyzed hydrolysis and cascade amplification.* Biosensors and Bioelectronics, 2017. **94**: p. 651-656.

44. Yang, Y., et al., *A pH-responsive bioassay for paper-based diagnosis of exosomes via mussel-inspired surface chemistry.* Talanta, 2019. **192**: p. 325-330.

45. Ali, J., et al., *Near infrared spectroscopy and imaging to probe differences in water content in normal and cancer human prostate tissues.* Technology in cancer research & treatment, 2004. **3**(5): p. 491-497.

46. Liu, K.-Z., et al., *Quantitative determination of serum LDL cholesterol by near-infrared spectroscopy.* Vibrational spectroscopy, 2005. **38**(1-2): p. 203-208.

47. Petiot, E., et al., *In situ quantification of microcarrier animal cell cultures using near-infrared spectroscopy.* Process Biochemistry, 2010. **45**(8): p. 1427-1431.

48. Mishra, P., et al., *Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction.* Talanta, 2021. **229**: p. 122303.

49. Martens, H. and T. Naes, *Multivariate calibration*. 1992: John Wiley & Sons.

50. Escandar, G.M., et al., *A review of multivariate calibration methods applied to biomedical analysis.* Microchemical Journal, 2006. **82**(1): p. 29-42.

51. Wang, A. and L. Xie, *Technology using near infrared spectroscopic and multivariate analysis to determine the soluble solids content of citrus fruit.* Journal of Food Engineering, 2014. **143**: p. 17-24.

52. Bogomolov, A., et al., *Quantitative determination of fat and total protein in milk based on visible light scatter.* Food Chemistry, 2012. **134**(1): p. 412-418.

53. Tsenkova, R., et al., *Somatic cell count determination in cow's milk by near-infrared spectroscopy: A new diagnostic tool.* Journal of animal science, 2001. **79**(10): p. 2550-2557.

54. Bogomolov, A. and A. Melenteva, *Scatter-based quantitative spectroscopic analysis of milk fat and total protein in the region 400–1100 nm in the presence of fat globule size variability.* Chemometrics and Intelligent Laboratory Systems, 2013. **126**: p. 129-139.

55. Bogomolov, A., et al., *Reference-free spectroscopic determination of fat and protein in milk in the visible and near infrared region below 1000 nm using spatially resolved diffuse reflectance fiber probe.* Talanta, 2017. **167**: p. 563-572.

56. Bogomolov, A., A. Melenteva, and D.J. Dahm, *Fat globule size effect on visible and shortwave near infrared spectra of milk.* Journal of Near Infrared Spectroscopy, 2013. **21**(5): p. 435-440.

57. Surkova, A., et al., *Towards an optical multisensor system for dairy: Global calibration for fat analysis in homogenized milk.* Microchemical Journal, 2019. **149**: p. 104012.

58. Dahm, D.J., *Explaining some light scattering properties of milk using representative layer theory.* Journal of Near Infrared Spectroscopy, 2013. **21**(5): p. 323-339.

59. Tsenkova, R., *Aquaphotomics: dynamic spectroscopy of aqueous and biological systems describes peculiarities of water.* Journal of Near Infrared Spectroscopy, 2009. **17**(6): p. 303-313.

60. Kojić, D., et al., *Water confined in the local field of ions.* ChemPhysChem, 2014. **15**(18): p. 4077-4086.

61. Muncan, J. and R. Tsenkova, *Aquaphotomics—From innovative knowledge to integrative platform in science and technology.* Molecules, 2019. **24**(15): p. 2742.

62. Wypych, G., *Handbook of solvents*. 2001, Toronto; New York: ChemTec.

63. Day, C.-P., et al., *Recognition of observer effect is required for rigor and reproducibility of preclinical animal studies.* Cancer Cell, 2022. **40**(3): p. 231-232.