



**HAL**  
open science

## Interest of locally weighted regression to overcome nonlinear effects during in situ NIR monitoring of CHO cell culture parameters and antibody glycosylation

Daniel Zavala-ortiz, Bruno Ebel, Meng-yao Li, Dulce Ma. Barradas-dermitz, Patricia Hayward-jones, Maria Aguilar-uscanga, Annie Marc, Emmanuel Guedon

### ► To cite this version:

Daniel Zavala-ortiz, Bruno Ebel, Meng-yao Li, Dulce Ma. Barradas-dermitz, Patricia Hayward-jones, et al.. Interest of locally weighted regression to overcome nonlinear effects during in situ NIR monitoring of CHO cell culture parameters and antibody glycosylation. *Biotechnology Progress*, 2020, 36 (e2924), pp.1-10. 10.1002/btpr.2924 . hal-02333228

**HAL Id: hal-02333228**

**<https://hal.univ-lorraine.fr/hal-02333228>**

Submitted on 2 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## RESEARCH ARTICLE

# Interest of locally weighted regression to overcome nonlinear effects during in situ NIR monitoring of CHO cell culture parameters and antibody glycosylation

Daniel A. Zavala-Ortiz<sup>1,2</sup>  | Bruno Ebel<sup>1</sup>  | Meng-Yao Li<sup>1</sup> |  
Dulce Ma. Barradas-Dermitz<sup>2</sup> | Patricia M. Hayward-Jones<sup>2</sup> |  
Maria G. Aguilar-Uscanga<sup>2</sup>  | Annie Marc<sup>1</sup>  | Emmanuel Guedon<sup>1</sup>

<sup>1</sup>Laboratoire Réactions et Génie des Procédés, Université de Lorraine, CNRS, Vandœuvre-lès-Nancy, France

<sup>2</sup>Tecnológico Nacional de México, Instituto Tecnológico de Veracruz, Veracruz, Veracruz, Mexico

## Correspondence

Bruno Ebel, Laboratoire Réactions et Génie des Procédés, CNRS-Lorraine University, UMR 7274, ENSAIA, TSA 40602, 54518 Vandœuvre-lès-Nancy, France.  
Email: bruno.ebel@univ-lorraine.fr

## Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1002/btpr.2924>.

## Abstract

Animal cell culture processes have become the standard platform to produce therapeutic proteins such as recombinant monoclonal antibodies (mAb). Since the mAb quality could be subject to significant changes depending on manufacturing process conditions, real time monitoring and control systems are required to ensure mAb specifications mainly glycosylation and patient safety. Up to now, real time monitoring glycosylation of proteins has received scarce attention. In this article, the use of near infrared (NIR) to monitor mAb glycosylation has been reported for the first time. Whereas monitoring models are mainly constructed using linear partial least squares regressions (PLSR), evidences presented in this study indicate nonlinearity relationship between in situ captured spectra and compound concentrations, compromising the PLSR performances. A novel and simple approach was proposed to fit nonlinearity using the locally weighted regression (LWR). The LWR models were found to be more appropriate for handling information contained in spectra so that real time monitoring of cultures were accurately performed. Moreover, for the first time, the LWR calibration models allowed mAb glycosylation to be monitored, in a real time manner, by using in situ NIR spectroscopy. These results represent a further step toward developing active-control feedback of animal cell processes, particularly for ensuring properties of biologics.

## KEYWORDS

in situ near infrared spectroscopy, locally weighted regression-LWR, monoclonal antibody-mAb glycosylation, partial least squares regression-PLSR, real time monitoring

## 1 | INTRODUCTION

The production of biologicals, especially recombinant monoclonal antibodies (mAb), remains a challenge due to the structural complexity of these molecules and their sensitivity to changes in the manufacturing process. That is why strict quality control systems are required to ensure

mAb specifications and patient safety. For that, regulatory agencies proposed the quality by design strategy,<sup>1</sup> which is rendered possible through the process analytical technology (PAT) approach.<sup>2-4</sup> The main objective is to real time monitor the concentrations of some process parameters, such as viable cell, nutrient, and metabolite concentrations, whose variability may have an impact on mAb quality attributes.<sup>5</sup>

As one of the main quality attributes, glycosylation pattern confers chemical and therapeutic properties to mAb (serum half-life, immunogenicity, antibody-dependent cellular cytotoxicity, and complement-dependent cytotoxicity).<sup>6-8</sup> Therefore, its control is essential to ensure efficacy of the product and safety for patients. However, glycosylation monitoring is usually performed at the end of the culture process because its analysis is time- and labor-consuming.<sup>9</sup> Even in advanced production processes supporting the PAT initiative, glycosylation-monitoring still requires some sample handling through the semiautomatic off-line analysis.<sup>10-12</sup> Moreover, delays due to off-line analysis may also compromise real time control of the process. A new challenging objective for PAT is thus to control mAb glycosylation as well as cell metabolism using online spectroscopy.<sup>13-16</sup> Consequently accurate monitoring models must be developed so that advanced active feedback control systems for controlling processes could become feasible.

In recent decades, vibrational near infrared (NIR) and Raman spectroscopies in combination with multivariate data analysis have been proven to be promising tools for monitoring cell culture process parameters.<sup>16-19</sup> The most widely used multivariate method for developing calibration models from spectroscopic data are partial least squares regression (PLSR).<sup>20</sup> PLSR maps linearly spectroscopic spectrum into a low-dimensional space of coordinates called latent variables (LV), which are employed to generate the regression or calibration equations using only linear combinations.<sup>21</sup> In this context, in situ spectroscopic monitoring has been claimed as an ideal analysis method since it provides real time multicomponent information directly without sample treatments, thus avoiding contamination risks and perturbation of compound properties.<sup>22</sup> Nevertheless, this in-line implementation remains a challenge because analyses can be subject to perturbations due to the dynamics of the cell culture. Indeed, scattering compounds are generated during cultures, such as cells, cellular debris, or mAb aggregates. Their accumulation in the bioreactor may induce modifications of the scattered light reaching the detector, causing changes in apparent absorbance, and thus resulting in nonlinear spectra slopes changes.<sup>23</sup> These effects, which induce or increase the nonlinear relationships between spectra and compounds, may limit the capabilities of classical linear regression methods, particularly PLSR.

Off-line spectroscopy and PLSR have been successfully used to perform monitoring of glycoproteins, particularly Raman spectroscopy.<sup>24</sup> However, clarified samples with limited scattering compounds, have been usually required to develop accurate calibration models for glycoproteins.<sup>25</sup> Such current off-line approaches cause monitoring delays that could compromise real time control of the process. As far as can be ascertained, while only one study has reported in situ mAb quality monitoring by Raman spectroscopy and PLSR,<sup>26</sup> such kind of monitoring using NIR spectroscopy has not been reported yet. This could be due to the complexity of the NIR spectra in scattering systems and the need of chemometric approaches adapted to such systems.<sup>27,28</sup> Taking into account the limitations of linear regression methods, several other methods and techniques are available to extract the information from spectra and overcome nonlinearity during the calibration process. They include deletion of nonlinear spectra variables, addition of nonlinear extra-terms to calibration equations, or use of other regression

methods.<sup>29-32</sup> The studies dealing with animal cell culture monitoring have been mostly restricted to selection of the best spectra variables using linear regressions. However, such approaches fail to properly address potential nonlinear relationships between spectra variabilities and compound concentration changes during cell cultures.

One way to solve nonlinear behavior using the widely known linear regression methods is to perform regression locally, such as the locally weighted regression (LWR) method. This method aims to model nonlinearity using several local linear regressions.<sup>33</sup> LWR requires defining the local area that contains the sample to be predicted. Then, each point in this local region is weighted according to its distance from the sample. Thus, in addition to the generation of the regression function for each local area, other parameters have to be determined in order to determine the local area and the point weights.<sup>34</sup> Consequently, a local linear calibration equation has to be generated using local data, each time a prediction is required. In contrast to global regressions that treat all the regression surface at the same time, as either linear or nonlinear, LWR allows modeling the nonlinear regions without compromising linear region predictions. This approach seems particularly adequate for animal cell culture processes, in which linear and nonlinear behavior may arise differently during exponential, stationary or declining growth phases of cell cultures, especially because of light scattering by cells, rendering parameter monitoring difficult.

Therefore, the aim of this article was to propose an adequate regression method for in-line monitoring of CHO cell cultures in a bioreactor, including real time analysis of mAb glycosylation site occupancy, by using in situ NIR spectroscopy. Firstly, it assessed the capacity of PLSR method to handle eventual nonlinear behavior during animal cell cultures. Secondly, the LWR method was proven capable of dealing with PLSR limitations so that in situ monitoring of various culture parameters as well as mAb concentrations and glycosylation patterns throughout batch cultures was possible.

## 2 | MATERIALS AND METHODS

### 2.1 | Cell cultures and NIR spectra acquisition

The genetically modified DG44 CHO cell line (CHO M250-9), producing a human anti-Rhesus D mAb, was kindly provided by Bioprocessing Technology Institute (Singapore). The culture medium was a protein-free medium mixture consisting of PF-CHO (HyClone) and CD-CHO (Thermo Fisher Scientific) in a 1:1 volume ratio, supplemented with 4 mM L-glutamine (Sigma Aldrich) and 0.1% pluronic F-68 (Sigma Aldrich). Cell cultures were performed in 2 L benchtop bioreactors (Pierre Guérin) with 1.5 L working volume. The set points for all cultures were 37°C, 50% dissolved oxygen, pH 7.2 and 90 rpm. Six cultures were performed in order to obtain off-line measurements, which cover bioprocess variability (three batches, two feed-harvest, and one batch with glucose spiking). In situ spectral scanning of bioprocess culture media was carried out using a NIR transreflectance probe with 1 mm path length (Precision Sensing Devices), connected to the Antaris II spectrometer (Thermo Fisher Scientific). Each NIR

spectrum corresponded to an average of 128 scans with an  $8\text{ cm}^{-1}$  resolution from  $4,000$  to  $10,000\text{ cm}^{-1}$  wave number (i.e.,  $2,500\text{--}1,000\text{ nm}$  wavelength).

## 2.2 | Off-line measurements

Viable cell density (VCD) was measured using the Vi-Cell XR™ cell counter (Beckman Coulter). Off-line concentrations of glucose, lactate, glutamine, and mAb were determined with enzymatic kits using the automated photometric analyzer Gallery™ (Thermo Fisher Scientific). Using UHPLC-MS as previously described,<sup>26</sup> mAb glycosylation heterogeneity was elucidated.

## 2.3 | Development of NIR calibration models with chemometric methods

Main calibration set, which consisted in data from batch, repeated-batch and glucose-spiked cultures, was randomly partitioned into calibration (80%) and validation set (20%). The calibration set involved 134 samples, while the validation set 34. The calibration set was used for cross-validation. Since validation set was not used for calibration processes, it was used to infer the performance of models in cultures completely unknown by the models. A maximum of four observations were deleted when they were identified as influential outliers (Hotelling's T-square method,  $p = .95$ ).<sup>28</sup> Firstly, in order to generate the PLSR calibration models for concentration of viable cells, glucose, lactate, glutamine, mAb and nonglycosylated mAb, special attention was given to NIR spectra preprocessing. Selection of spectra preprocessing methods was an exhaustive qualitative process to determine and mitigate additive-multiplicative effects and wavelength-dependent baseline variation.<sup>27</sup> Not less than 30 pretreatments and their combinations were compared based on model performance evaluated by the root mean squared error of cross-validation (RMSECV) and the root mean squared error of prediction (RMSEP). RMSECV and RMSEP allow a direct measure of accuracy using calibration data and independent data, respectively. A model with lower RMSECV or RMSEP is considered more accurate. The relative error (RE) was also used to evaluate the models. RE is the relationship of RMSECV or RMSEP with the maximum concentration of a compound during the calibration process. It is used as a contextualized error of a model so that comparison of accuracy is more meaningful to expected measures during real time calculations. The particular spectra pretreatment that led to higher accuracy for a compound PLSR model was given in Table 1.

Subsequently, calibration models were generated using the LWR method. An optimization process was carried out to determine the local areas for regression in terms of number of local points (LPs) and the level of spectra compression in term of principal components (PC).<sup>28</sup> LP is the number of nearest calibration samples in the principal component space to be used for a particular local regression, and it represents a good measure of nonlinearity. A model with few LPs suggests the presence of strong nonlinearity, since, more local regressions based on only few samples are needed to properly outline the nonlinear relationship between spectra and compounds. As PC is a linear spectrum mapping procedure, it will require more PC to fit a nonlinear relationship, then a large number of PC used in a

**TABLE 1** Spectral preprocessing used for PLSR and LWR models

|              | PLS models  | LWR models                                     |
|--------------|---|--|
| Viable cells | First derivative (1, 2, 15), abs, smoothing (6, 11) | MSC (mean)                                     |
| Glutamine    | Detrend   | Detrend, SNV                                   |
| Glucose      | SNV, First derivative (1, 15), abs                  | MSC (mean), First derivative (1, 9)            |
| Lactate      | Detrend, SNV  | Detrend  |
| mAb          | MSC (mean), Detrend, second derivative (2, 21), abs | MSC (mean), Detrend                            |
| NG-mAb       | Detrend, MSC (mean), second derivative (2, 15)      | Detrend, MSC (mean), second derivative (2, 15) |

Note: Values for derivatives are filter width, polynomial order, derivative order.

Abbreviations: MSC: Multiple scatter correction; SNV, standard normal variate; abs, absolute value; OSC, orthogonal scatter correction.

LWR model also suggest the presence of nonlinearity. Parallel to a former optimization, a spectrum pretreatment determination approach as used for PLSR was also performed. The particular spectrum pretreatment that led to higher accuracy for a compound LWR model is given in Table 1.

As established for analytical procedures, linearity means the ability of a model to obtain results directly and linearly proportional to actual concentration within a compound concentration range.<sup>35</sup> Then non-linearity and incapacity of models to handle it, can be detected by a systematic deviation of residuals from the zero line, usually with a curve tendency.<sup>28</sup> In order to properly detect nonlinear behavior during calibration, residual plots (off-line measured values against difference between off-line measurements and in-line calculated values) were analyzed as suggested elsewhere.<sup>36,37</sup> The presence of nonlinear relationships was confirmed by visual inspection of these plots. If models are capable to handle spectra linearly in relation to concentration, residuals have a random distribution around the zero line. On the contrary, nonlinearity is detected by a correlation of residuals, usually in a curve profile. Once nonlinear behavior is visually found, its management by models is statistically taken into account by evaluation of residual correlation by the Durbin-Watson test.<sup>38</sup> This test evaluates the null hypothesis that there is no correlation between successive residuals (random distribution of residuals around the zero line), which is likely the case if the model estimates concentration from spectra linearly to actual concentration. Rejection of the null hypothesis indicates that correlation exists and that the model leads systematically to misestimation of concentrations because of nonlinear relationship existing between spectra and concentration. Multivariate calibration models and statistical analysis were carried out using PLS-Toolbox 8.2.1 (Eigenvector Research Inc.) and Statistics, and Machine Learning Toolbox R2016a in MATLAB® environment (MathWorks Inc.).

## 2.4 | In situ monitoring of cell cultures

To evaluate the predictive capacity of the models, NIR spectra were automatically acquired in situ every 20 min throughout CHO cell

cultures. The optimized NIR calibration models were then used to perform real time calculation of the concentrations of main culture medium components (viable cells, glucose, lactate, glutamine, and mAb) as well as of the nonglycosylated mAb.

### 3 | RESULTS AND DISCUSSION

NIR spectroscopy has been proven as useful tool to monitor animal cell cultures under the PAT approach. However, the capabilities of NIR models to detect glycosylation of produced proteins in situ remain unknown. Moreover, in situ NIR spectroscopy monitoring still requires to be improved so that more accurate estimations of compound concentration could be used to also monitor cell metabolism. As far as can be ascertained, only PLSR has been used to monitor cell cultures in situ based on NIR spectroscopy. In this study, PLSR and LWR models were generated and compared to estimate CHO viable cells, glucose, lactate, and glutamine concentrations. In addition, models for total mAb and nonglycosylated mAb (NG-mAb) concentrations were also generated with the aim of monitoring the quality of mAbs produced during a process.

#### 3.1 | Development and analysis of NIR models based on PLSR and LWR methods

##### 3.1.1 | Development of PLSR models for monitoring of animal cell culture process

Evaluation of PLSR models comprised different statistical estimators of accuracy, precision and linearity. The coefficient of determination ( $R^2$ ) is a parameter used to evaluate how the model explains

concentration variability based on spectrum variability. Coefficients with values close to one indicate that models relate spectra variability to concentration, while values close to zero suggest no relationship between spectra and concentration. In this context, PLSR was capable of relating spectra variabilities to concentrations for viable cells and glucose, as shown by the high  $R^2$  values in Table 2. Consequently, PLSR models achieved high accuracy during the calibration and prediction steps as indicated by RMSECV and RMSEP, respectively. However, PLSR showed limited capacity to relate spectra variabilities to lactate and glutamine concentrations as indicated by the low  $R^2$  values. PLSR models for such compounds resulted in REs above 10%, usually perceived as the maximum tolerated value for NIR calibration models.<sup>39</sup> This phenomenon was also observed for the mAb and NG-mAb PLSR models, with approximately 10 and 20% of REs of prediction ( $R_{PRE}$ ), respectively, indicating that PLSR is not the appropriate multivariate technique for monitoring mAb concentration and quality during animal cell culture processes.

##### 3.1.2 | Nonlinear relationships between cell culture parameters and spectra

The way a model fits data induces some specific structure distribution of the residuals. If a linear model such as PLSR fits a curve relationship between spectra and compound concentrations, residuals will have a systematic deviation. In contrast, if the linear model fits a linear relationship, the residuals would scatter randomly from the zero line. This criterion was then used to graphically detect nonlinear relationships and adequateness of PLSR (Figure 1). A clear hyperbolic profile was

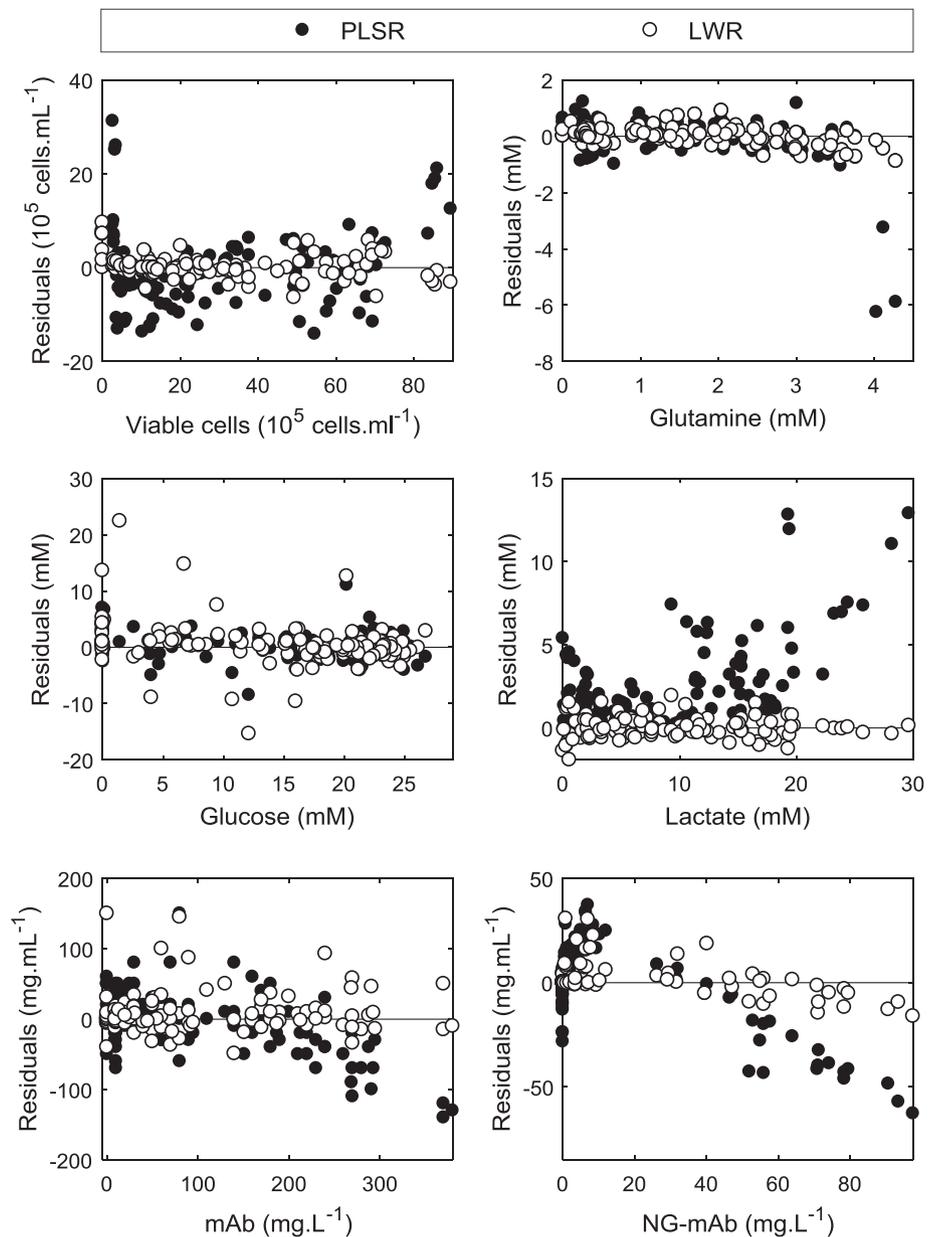
|      |                  | Viable cells | Glucose | Lactate | Glutamine | mAb   | NG-mAb |
|------|------------------|--------------|---------|---------|-----------|-------|--------|
| PLSR | RMSECV           | 6.15         | 1.94    | 3.22    | 0.44      | 35    | 19     |
|      | RMSEP            | 8.34         | 2.51    | 4.38    | 0.50      | 53    | 18     |
|      | $R^2_{CV}$       | 0.95         | 0.95    | 0.78    | 0.85      | 0.84  | 0.39   |
|      | $R^2_{PRE}$      | 0.95         | 0.93    | 0.67    | 0.81      | 0.83  | 0.50   |
|      | R. $E_{CV}$ (%)  | 6.47         | 7.19    | 10.73   | 10.23     | 9.21  | 19.00  |
|      | R. $E_{PRE}$ (%) | 8.78         | 9.30    | 14.60   | 11.63     | 13.95 | 18.00  |
|      | LV               | 5            | 6       | 5       | 5         | 5     | 3      |
| LWR  | RMSECV           | 4.97         | 3.91    | 1.78    | 0.29      | 37    | 7.1    |
|      | RMSEP            | 5.50         | 3.62    | 2.48    | 0.42      | 45    | 11     |
|      | $R^2_{CV}$       | 0.97         | 0.81    | 0.87    | 0.93      | 0.86  | 0.92   |
|      | $R^2_{PRE}$      | 0.97         | 0.86    | 0.82    | 0.87      | 0.84  | 0.80   |
|      | R. $E_{CV}$ (%)  | 5.23         | 14.48   | 5.93    | 6.74      | 9.74  | 7.00   |
|      | R. $E_{PRE}$ (%) | 5.89         | 13.41   | 8.27    | 9.77      | 11.84 | 11.00  |
|      | LP-PC            | 13-6         | 5-4     | 13-3    | 7-3       | 21-4  | 15-15  |

**TABLE 2** Statistical analysis of calibration models using PLSR or LWR method

Note: Units for RMSECV and RMSEP are the same:  $(\text{cells} \cdot \text{ml}^{-1}) \times 10^5$  for viable cells,  $\text{mg} \cdot \text{L}^{-1}$  for mAb and NG-mAb, and mM for glucose, lactate, and glutamine.

Abbreviations: LP, number of local points; LV, number of latent variables; NG-mAb, non-glycosylated mAb; PC, number of principal components; R.  $E_{CV}$ , R.  $E_{PRE}$ , relative errors (in relation to maximum variable concentration) for cross validation (CV) or prediction (PRE);  $R^2_{CV}$ ,  $R^2_{PRE}$ , correlation coefficients of cross-validation or prediction; RMSECV, root mean square error of cross-validation; RMSEP, root mean square error of prediction.

**FIGURE 1** Detection of nonlinearity by inspection of residual distribution within the concentration range tested during calibration for viable cells, glutamine, glucose, lactate, mAb and NG-mAb concentration (off-line concentration against difference between off-line and estimated concentration by models). Residuals were obtained using PLSR (●) or LWR (○) regression method. LWR, locally weighted regression; PLSR, partial least squares regressions



observed for NG-mAb residuals, indicating that a strong nonlinear relationship occurred between spectra and NG-mAb concentrations for the whole concentration range tested. In such a case, a linear PLSR model is inadequate and therefore nonlinear regression approaches are required to generate proper calibration models. This nonlinear behavior was also observed for mAb to a lesser extent. In addition, a flattened parabola profile was observed for viable cells and lactate concentrations, while a curve tendency was detected for glutamine at concentrations over 3 mM. In these cases, PLSR is considered as adequate only for the concentration range where a linear relationship between actual and estimated compound concentrations is observed, corresponding to a random distribution of the residuals.

Nonlinearity was then statistically analyzed by the use of the Durbin–Watson test. The test provides a *d*-value related to the nature of the residual distribution, which is then used to evaluate correlation

taking into account the nature of the calibration set. Results from this analysis revealed correlation of residuals for all PLSR models (Table 3), which particularly indicated a strong nonlinearity for lactate and NG-mAb. These results suggested, from a statistical point of view, the inability of the PLSR to obtain results directly and linearly proportional to actual concentrations. Other widely used linear approaches, such as principal component regression, has also been assessed with similar results (data not shown). This is in agreement with a former study reporting similar or even lower performances of these linear approaches.<sup>40</sup> Therefore, the novel implementation of other regression methods considering linearity analysis must be considered.

The presence of nonlinear behavior, which has limited the performance of PLSR, may be explained by the complexity of the cell culture medium used during the process. As NIR spectra contain both physical and chemical information of the samples, it is likely that nonlinearity

**TABLE 3** Results of Durbin–Watson analysis

| Compound     | PLSR            |            | LWR             |            |
|--------------|-----------------|------------|-----------------|------------|
|              | <i>d</i> -value | Conclusion | <i>d</i> -value | Conclusion |
| Viable cells | 0.73            | Nonlinear  | 1.78            | Linear     |
| Glutamine    | 0.99            | Nonlinear  | 1.71            | Nonlinear  |
| Glucose      | 1.65            | Nonlinear  | 1.75            | Linear     |
| Lactate      | 0.40            | Nonlinear  | 1.92            | Linear     |
| mAb          | 0.89            | Nonlinear  | 1.86            | Linear     |
| NG-mAb       | 0.46            | Nonlinear  | 0.8             | Nonlinear  |

Note: The critical *d* values for Durbin–Watson at  $\alpha = .05$  are  $d_L = 1.72$  and  $d_U = 1.74$ . If  $d < d_L$  the null hypothesis is rejected, the presence of correlation in residuals is indicated (nonlinearity). If  $d > d_U$  the null hypothesis cannot be rejected, the correlation in residuals is considered to be negligible (no nonlinearity). If  $d_L < d < d_U$  the test is inconclusive.

resulted from a wide variety of phenomena, such as variations of light diffusion profiles during cultures.<sup>41</sup> Chemical phenomena are mainly related to changes in the interaction of several absorbing functional groups, which may lead to shifts of absorption bands or to effects such as the Fermi and Darling–Dennison resonances.<sup>42</sup> Such resonance phenomena might require management of nonlinearity relationship to properly extract the information within spectra.<sup>43</sup> Therefore, the use of the LWR method has been evaluated with the aim of overcoming the limitations of the linear PLSR method.

### 3.1.3 | Development of LWR models to handle nonlinear relationships and comparison with PLSR models

The development of LWR models firstly comprised an optimization process to select the size of local areas and the number of PC. If the size of local areas is small, more local regressions must be launched to fit global nonlinearities, which also influence the way information within spectra should be handled. This is particularly related to the number of PC required to perform local regressions. Consequently, a compromise between local areas size, in terms of LPs and PC had to be found to avoid LWR overfitting. The final structure of LWR models is given in Table 2. The size of local areas in term of number of local points varied from 5 to 21 for the different compounds, which represented ~10% of the calibration set and depicted a strong nonlinear behavior. This nonlinearity is likely attributed to the dynamics of the culture process since the locations of the LPs used for the local regressions were mainly determined as a function of culture progression. On the other hand, a high number of PC likely depicts a nonlinearity caused by an inherent nonlinear relationship between spectra and concentration. Once the global nonlinearity was broken, relatively few PC were required to build the local regressions for viable cells, lactate, glutamine, and mAb concentration. However, for NG-mAb and glucose, a relatively high number of PC, depending on the number of local points, was required. The LWR models related concentrations of lactate and glutamine with spectral variability more efficiently, as

shown by the higher  $R^2$  in contrast with results obtained using PLSR (Table 2). This enhanced management of spectral variability by the LWR method resulted in a reduction of RMSECV and RMSEP values, which corresponded to decreased REs of about 35% for lactate and glutamine. Such reductions allowed concentration estimates with R.E lower than 10%. The LWR method also enhanced the accuracy for a viable cell model, resulting in a reduction of R.E of ~30%. LWR displayed a similar performance to PLSR for mAb concentration whereas PLSR was higher for glucose (Table 2).

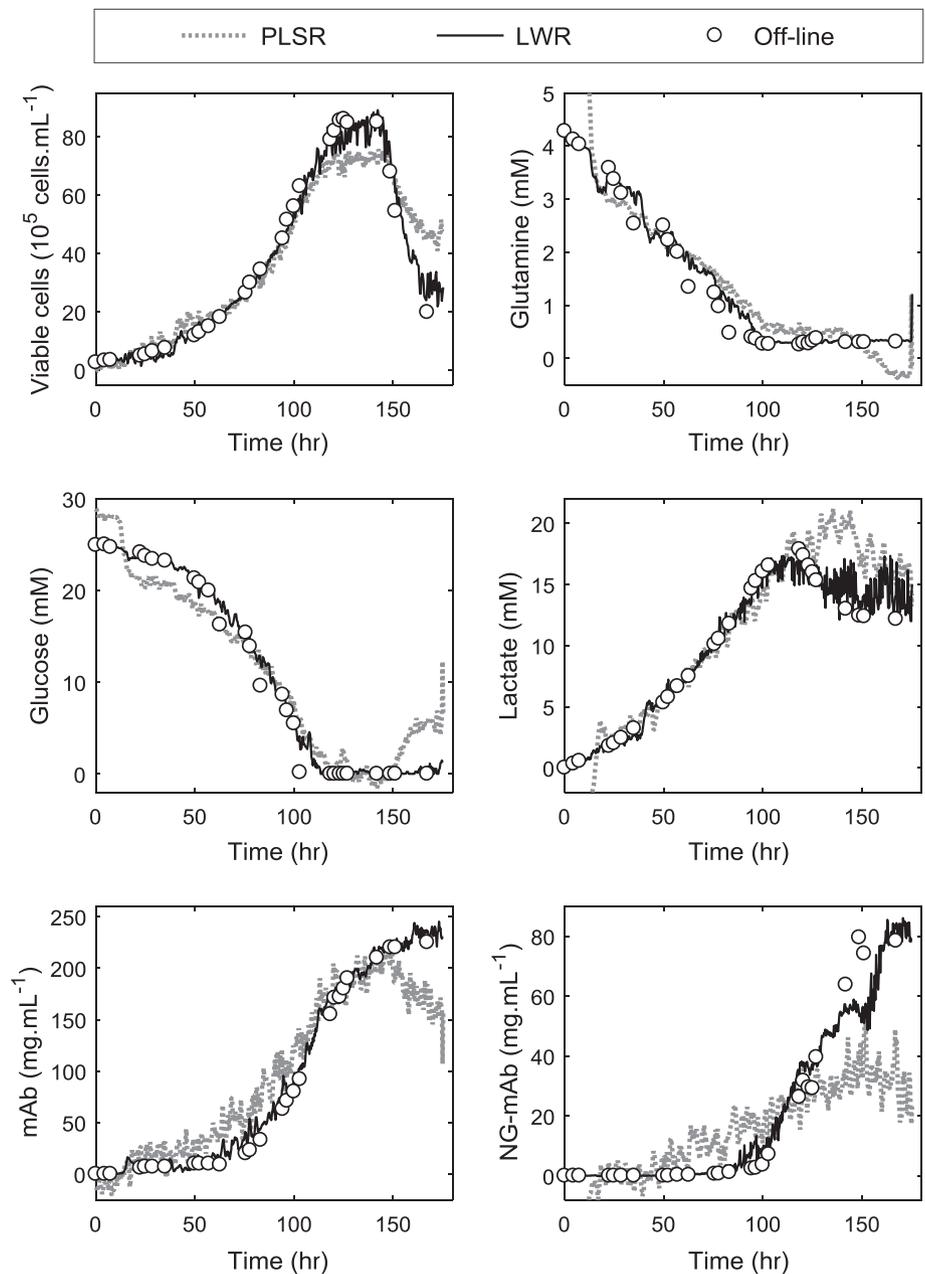
A remarkable characteristic of the LWR method was its capability to handle the strong nonlinearity behavior previously detected between spectra and concentration, as shown by the higher  $R^2$ , particularly during cross-validation, in comparison to PLSR (Table 2). Analysis of residual plots (Figure 1) confirmed that in general, LWR not only enhanced accuracy, but also drastically limited the effects of nonlinearity, particularly for NG-mAb. This was statistically confirmed by the Durbin–Watson test (Table 3), which indicated that, excepting NG-mAb and glutamine models, all LWR models properly handled nonlinearity and estimated compound concentration linearly to actual concentrations. In this context, the LWR method allowed the development of a calibration model with REs of ~9%, opening the possibility to monitor the quality of mAb in terms of glycosylation site occupancy. As a general rule, the LWR method appeared to be the most appropriate model method for the majority of compounds, particularly NG-mAb.

### 3.2 | Real time monitoring of animal cell culture processes

Performances of prediction models based on both PLSR and LWR methods were evaluated during CHO cell culture processes producing mAb in a discontinuous mode. In-line monitoring of viable cells, glutamine, glucose, lactate, mAb, and NG-mAb concentrations was carried out using NIR spectroscopy (Figure 2). In-line predictions were compared with off-line measurements to verify model accuracy. In all cases, the LWR method showed enhanced performance during real time monitoring.

Contrary to what was expected on the basis of a relative high  $R^2$  (Table 1) the PLSR model was not able to monitor viable cell concentrations efficiently from ~100 hr of culture. From this moment, the model was unable to estimate viable cell concentrations from NIR spectra. This unsatisfactory result is in agreement with former reports that pointed out difficulties to model viable cell concentration from NIR spectra due to scattering effects.<sup>44,45</sup> For example, only limited results of PLSR models of viable cell concentrations were obtained despite the high number of 10 LV required.<sup>44</sup> In addition, even with promising PLSR models established for viable cell concentrations, low precision and accuracy appeared once viable cell densities exceeded  $80.10^5$  cells  $\text{ml}^{-1}$ .<sup>45</sup> The lack of accuracy of the results could be attributed to the nonlinear relationship between spectra and viable cell concentrations. This was previously observed by the right segment of the parabola in the residual plot for viable cell concentrations, which corresponds to concentration values over  $75.10^5$  cells  $\text{ml}^{-1}$  (Figure 1).

**FIGURE 2** Real time monitoring of a CHO cell culture in batch reactor by in situ NIR spectroscopy. Comparison of in-line prediction by models using PLSR (●) or LWR (—) or regression methods with experimental off-line results (○). LWR, locally weighted regression; PLSR, partial least squares regressions



Thereby, a more reliable comparison between PLSR and LWR models, with similar  $R^2$  and REs, necessarily requires this linearity analysis to be performed in order to better evaluate the calibration process for in-line monitoring of animal cell cultures. A similar phenomenon was observed for the glucose PLSR model, which was even better than LWR during calibration. However, during prediction, LWR accurately monitored glucose concentration during the whole culture while the PLSR model misestimated the concentrations at the beginning and end of the culture. This behavior was also observed for glutamine and lactate (Figure 2). The analysis of spectra location in the PLSR space revealed that for the first 10 hr of culture they were totally outside the calibration space at 95% confidence limit (data not shown). Thus, the observed mis-prediction at the beginning of the culture could be attributed to an inappropriate extrapolation. On the contrary, the

LWR method was shown to be more robust during this phase since culture spectra were always incorporated into the local PLSR space in order to perform a regression. The misestimations of lactate, glucose, and glutamine concentrations after 120 hr of the culture by PLSR model could also be partially explained by the nonlinearity phenomenon as formerly observed in Figure 1. They could also be attributed to the lack of precision or to the presence of noise since even the LWR method showed prediction errors, although weaker, after 120 hr. In addition, after 120 hr, the culture is characterized by a decrease in cell viability, resulting in the release of intracellular metabolites. Such metabolites could interact with the monitored molecules causing discrepancies in NIR spectra. This lack of precision of PLSR models at the end of cultures has already been reported, probably due to unknown components within some batches, while some other batches showed

good predictions.<sup>44</sup> Most of the studies, which present the use of PLSR as an efficient method to predict animal cell culture parameters using NIR spectroscopy, do not consider concentration ranges as broad as those reported here, neither the cell death-phase for the process monitoring.<sup>18,22,46</sup> Some other authors have overcome non-linearity behavior of NIR monitoring models by using at-line analyses of clarified culture medium, as already described.<sup>47-49</sup> In this context, LWR has been proven as a promising method to perform in situ monitoring since it allowed calibration in a wide concentration frame, and also considered the nature of the culture medium.

### 3.3 | Real time monitoring of mAb concentration and quality during animal cell culture processes

Previously developed methods using PLSR and LWR were used to monitor total mAb concentration and nonglycosylated mAb. While the PLSR method showed a good trend for the evaluations of total mAb concentrations during the first days of cell culture, a reduced precision was observed from 140 hr. This resulting misestimation was likely due to the enrichment of NG-mAb within the total mAb molecules. Consequently, results indicated that even models apparently able to monitor properly total mAb concentration could be strongly influenced by changes in mAb properties. This justifies the use of other regression methods, as confirmed by the good performance of LWR prediction model of total-mAb (Figure 2).

The glycosylation pattern is a key quality parameter of mAb since it confers important properties such as ADCC or serum half-life.<sup>6-8</sup> Therefore its monitoring and control during process productions as proposed by the PAT initiative, is mandatory to ensure efficacy of mAb and safety of patients. Glycosylation analysis is time-consuming and usually performed by off-line approaches, which may induce a monitoring delay, and then compromise corrective action in order to maintain desired glycosylation properties. In this work, in situ NIR spectroscopy capability to real time monitor nonglycosylated mAb concentrations has been proven, provided the LWR method was used in place of the PLSR.

The performance of PLSR to monitor NG-mAb concentrations was first evaluated. As shown in Figure 2, the PLSR method completely failed to predict accurately NG-mAb concentrations throughout the culture, accordingly to low  $R^2$  values (Table 2). This result was expected because during calibration, a strong nonlinear relationship between residuals and NG-mAb concentrations was observed (Figure 1), indicating the need for a more efficient regression method to monitor NG-mAb concentrations by in situ NIR spectroscopy. Indeed, the use of the LWR method made it possible to reduce the prediction errors and to obtain a good monitoring of NG-mAb concentrations (Figure 2).

## 4 | CONCLUSIONS

In this study, experimental evidence of nonlinear parameter behavior in animal cell culture processes was provided. Consequently, the

widely used PLSR method was incapable of relating spectra with compound concentrations, indicating that such a widely used regression method is not always appropriate for the monitoring of animal cell culture processes. The novel use of the LWR method was shown to overcome PLSR limitations, which led to more accurate predictions of culture compound concentrations. Using NIR spectroscopy, the enhanced capability of LWR to handle nonlinearities permitted for the first time, the in situ monitoring of mAb glycosylation site occupancy. Overall, the results highlighted the fact that in situ NIR spectroscopy could have a broader potential as a PAT tool provided that effect of culture dynamics and nonlinearity be considered. In this context, NIR spectroscopy could be used to develop innovative spectroscopic calibration models so that effective control approaches to guarantee quality of antibodies could be implemented.

### ACKNOWLEDGMENTS

The authors acknowledge all national agencies that supported Daniel A. Zavala-Ortiz: the National Council of Science and Technology of Mexico for a PhD scholarship, the Veracruz Institute of Technology for a tuition fee scholarship, the French Ministry of Europe and Foreign Affairs for the Eiffel Excellence Scholarship. They also acknowledge the French National Agency for Research (ANR) for supporting the ProCell-In-Line project, as well as B.M. Wise from Eigenvector Research and C. Paris from Platform of Structural and Metabolomics Analyses (Nancy) for their kind support. D.A. Zavala-Ortiz particularly acknowledges the economic support offered by the quimioteraxol team members.

### CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

### ORCID

Daniel A. Zavala-Ortiz  <https://orcid.org/0000-0002-9572-4807>

Bruno Ebel  <https://orcid.org/0000-0003-3270-4881>

Maria G. Aguilar-Uscanga  <https://orcid.org/0000-0002-3875-7928>

Annie Marc  <https://orcid.org/0000-0003-1373-7119>

### REFERENCES

1. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. *Nat Biotechnol.* 2009;27(1):26-34.
2. Yu LX, Amidon G, Khan MA, et al. Understanding pharmaceutical quality by design. *AAPS J.* 2014;16(4):771-783.
3. Rathore AS, Bhambure R, Ghare V. Process analytical technology (PAT) for biopharmaceutical products. *Anal Bioanal Chem.* 2010;398(1):137-154.
4. Yu LX. Pharmaceutical quality by design: product and process development, understanding, and control. *Pharm Res.* 2008;25:781-791.
5. Teixeira AP, Oliveira R, Alves PM, Carrondo MJT. Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative. *Biotechnol Adv.* 2009;27(6):726-732.

6. Kayser V, Chennamsetty N, Voynov V, Forrer K, Helk B, Trout BL. Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnol J*. 2011;6(1):38-44.
7. Lingg N, Zhang P, Song Z, Bardor M. The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol J*. 2012;7(12):1462-1472.
8. del Val IJ, Kontoravdi C, Nagy JM. Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. *Biotechnol Prog*. 2010;26(6):1505-1527.
9. Huhn C, Selman MHJ, Ruhaak LR, Deelder AM, Wuhler M. IgG glycosylation analysis. *Proteomics*. 2009;9(4):882-913.
10. Alvarez M, Tremintin G, Wang J, et al. On-line characterization of monoclonal antibody variants by liquid chromatography-mass spectrometry operating in a two-dimensional format. *Anal Biochem*. 2011;419(1):17-25.
11. Wang Y, Feng P, Sosic Z, Zang L. Monitoring glycosylation profile and protein titer in cell culture samples using ZipChip CE-MS. *J Anal Bioanal Tech*. 2017;8(2):1-8.
12. Doherty M, Bones J, McLoughlin N, et al. An automated robotic platform for rapid profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture. *Anal Biochem*. 2013;442(1):10-18.
13. Zhang P, Woen S, Wang T, et al. Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discov Today*. 2016;21(5):740-765.
14. Hossler P. Protein glycosylation control in mammalian cell culture: past precedents and contemporary prospects. In: Hu W, Zeng A, eds. *Genomics and systems biology of mammalian cell culture*. Hamburg: Springer; 2011:187-219.
15. Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology*. 2009;19(9):936-949.
16. Berry BN, Dobrowsky TM, Timson RC, Kshirsagar R, Ryll T, Wiltberger K. Quick generation of Raman spectroscopy based in-process glucose control to influence biopharmaceutical protein product quality during mammalian cell culture. *Biotechnol Prog*. 2016;32(1):224-234.
17. Li M, Ebel B, Courtès F, Guedon E, Marc A. *In Situ* infrared spectroscopy as a PAT tool of great promise for real time monitoring of animal cell culture processes. *J Anal Bioanal Tech*. 2016;3(2):1065-1069.
18. Cervera AE, Petersen N, Lantz AE, Larsen A. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. *Biotechnol Prog*. 2009;25(6):1561-1581.
19. Abu-Absi NR, Kenty BM, Cuellar ME, et al. Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. *Biotechnol Bioeng*. 2010;108(5):1215-1221.
20. Lourenço ND, Lopes JA, Almeida CF, Sarraguça MC, Pinheiro HM. Bioreactor monitoring with spectroscopy and chemometrics: a review. *Anal Bioanal Chem*. 2012;404(4):1211-1237.
21. Höskuldsson A. PLS regression methods. *J Chemom*. 1988;2:211-228.
22. Arnold SA, Crowley J, Woods N, Harvey LM, McNeil B. *In-situ* near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. *Biotechnol Bioeng*. 2003;84(1):13-19.
23. Ge Z, Cavinato AG, Callis JB. Non-invasive spectroscopy for monitoring cell density in a fermentation process. *Anal Chem*. 1994;60:1354-1362.
24. Brewster VL, Ashton L, Goodacre R. Monitoring the glycosylation status of proteins using Raman spectroscopy. *Anal Chem*. 2011;83(15):6074-6081.
25. Li B, Raya BH, Leister KJ, Ryder AG. Performance monitoring of a mammalian cell based bioprocess using Raman spectroscopy. *Anal Chim Acta*. 2013;796:84-91.
26. Li M, Ebel B, Paris C, Chauchard F, Guedon E, Marc A. Real time monitoring of antibody glycosylation site occupancy by *in situ* Raman spectroscopy during bioreactor CHO cell cultures. *Biotechnol Prog*. 2018;34(2):486-493.
27. Huang J, Romero-Torres S, Moshgbar M. Practical considerations in data pre-treatment for NIR and Raman spectroscopy. *American Pharmace Rev*. 2010;13(6):116-127.
28. Naes T, Isaksson T, Fearn T, Davies T. *A user-friendly guide to multivariate calibration and classification*. Chichester: NIR Publications; 2002.
29. Mehmood T, Liland KH, Snipen L, Sæbø S. A review of variable selection methods in partial least squares regression. *Chemometr Intell Lab*. 2012;118:62-69.
30. Blanco M, Coello J, Iturriaga H, Maspoch S, Pagès J. Calibration in non-linear near infrared reflectance spectroscopy: a comparison of several methods. *Anal Chim Acta*. 1999;384(2):207-214.
31. Xiaobo Z, Jiewen Z, Povey MJ, Holmes M, Hanpin M. Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta*. 2010;667(1-2):14-32.
32. Verdú-Andrés J, Massart DL, Menardo C, Sterna C. Correction of non-linearities in spectroscopic multivariate calibration by using transformed original variables. Part II. Application to principal component regression. *Anal Chim Acta*. 1999;389(1-3):15-130.
33. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc*. 1988;83(403):596-610.
34. Centner V, Massart DL. Optimization in locally weighted regression. *Anal Chem*. 1998;70(19):4206-4211.
35. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. "Guideline for Industry: ICH-Q2A, Validation of Analytical Procedures". [https://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Quality/Q2\\_R1/Step4/Q2\\_R1\\_Guideline.pdf](https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1_Guideline.pdf). Accessed March 28, 2019.
36. Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J. *Handbook of chemometrics and qualimetrics: part A*. Amsterdam: Elsevier; 1997.
37. Centner V, de Noord OE, Massart DL. Detection of nonlinearity in multivariate calibration. *Anal Chim Acta*. 1998;376(2):153-168.
38. Howard MJ, Workman JJ. Chemometrics in spectroscopy - linearity in calibration: the Durbin-Watson statistic. *Spectroscopy*. 2005;20(3):34-40.
39. Workman JJ. Spectroscopy calibration basics. *Handbook of near-infrared analysis*. 3rd ed. Boca Raton: CRC Press; 2008:123-150.
40. Khajehsharifi H, Pourbasheer E, Tavallali H, Sarvi S, Sadeghi M. The comparison of partial least squares and principal component regression in simultaneous spectrophotometric determination of ascorbic acid, dopamine and uric acid in real samples. *Arab J Chem*. 2017;10: S3451-S3458.
41. Miller CE. Sources of non-linearity in near infrared methods. *NIR News*. 1993;4(6):3-5.
42. Siesler HW, Ozaki Y, Kawata S, Heise HM. *Near-infrared spectroscopy: principles, instruments, applications*. Weinheim: Wiley-VCH; 2008.
43. Agranovich VM, Kamchatrov AM. Fermi resonance nonlinear waves and solitons in organic superlattices. Boardman AD, Pavlov L, Tanev S. *advanced photonics with Secondorder optically nonlinear processes*. Sozopol: Springer Science+Business Media Oordrecht; 1998:251-275.
44. Henriques JG, Buziol S, Stocker E, Voogd A, Menezes JC. Monitoring mammalian cell cultivations for monoclonal antibody production using near-infrared spectroscopy. *Adv Biochem Engin/Biotechnol*. 2009;116:73-97.
45. Clavaud M, Roggo Y, Daeniken RV, Liebler A, Schwabe J. Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: prediction of multiple cultivation variables. *Talanta*. 2013;111:28-38.

46. Qiu J, Arnold MA, Murhammer DW. On-line near infrared bioreactor monitoring of cell density and concentrations of glucose and lactate during insect cell cultivation. *J Biotechnol.* 2014;173:106-111.
47. Rhiel MH, Cohen MB, Arnold MA, Murhammer DW. On-line monitoring of human prostate cancer cells in a perfusion rotating wall vessel by near-infrared spectroscopy. *Biotechnol Bioeng.* 2004;86(7):852-861.
48. Hakemeyer C, Strauss U, Werz S, Jose GE, Folque F, Menezes JC. At-line NIR spectroscopy as effective PAT monitoring technique in Mab cultivations during process development and manufacturing. *Talanta.* 2012;90:12-21.
49. Riley MR, Rhiel M, Zhou X, Arnold MA, Murhammer DW. Simultaneous measurement of glucose and glutamine in insect cell culture

media by near infrared spectroscopy. *Biotechnol Bioeng.* 2000;55(1):11-15.

**How to cite this article:** Zavala-Ortiz DA, Ebel B, Li M-Y, et al. Interest of locally weighted regression to overcome nonlinear effects during in situ NIR monitoring of CHO cell culture parameters and antibody glycosylation. *Biotechnol Progress.* 2020;36:e2924. <https://doi.org/10.1002/btpr.2924>