



**HAL**  
open science

## Support Vector and Locally Weighted regressions to monitor monoclonal antibody glycosylation during CHO cell culture processes, an enhanced alternative to Partial Least Squares regression

Daniel Zavala-Ortiz, Bruno Ebel, Mengyao Li, Dulce Ma. Barradas-dermitz, Patricia Hayward-Jones, Maria Aguilar-Uscanga, Annie Marc, Emmanuel Guedon

### ► To cite this version:

Daniel Zavala-Ortiz, Bruno Ebel, Mengyao Li, Dulce Ma. Barradas-dermitz, Patricia Hayward-Jones, et al.. Support Vector and Locally Weighted regressions to monitor monoclonal antibody glycosylation during CHO cell culture processes, an enhanced alternative to Partial Least Squares regression. Biochemical Engineering Journal, 2020, 154, pp.107457. 10.1016/j.bej.2019.107457 . hal-02376535

**HAL Id: hal-02376535**

**<https://hal.univ-lorraine.fr/hal-02376535>**

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



## Regular article

# Support Vector and Locally Weighted regressions to monitor monoclonal antibody glycosylation during CHO cell culture processes, an enhanced alternative to Partial Least Squares regression

Daniel Arturo Zavala-Ortiz<sup>a,b</sup>, Bruno Ebel<sup>a,\*</sup>, Meng-Yao Li<sup>a</sup>, Dulce María Barradas-Dermitz<sup>b</sup>, Patricia Margaret Hayward-Jones<sup>b</sup>, María Guadalupe Aguilar-Uscanga<sup>b</sup>, Annie Marc<sup>a</sup>, Emmanuel Guedon<sup>a,\*</sup>

<sup>a</sup> Laboratoire Réactions et Génie des Procédés, UMR 7274, CNRS-Université de Lorraine, 2 avenue Forêt de Haye, TSA 40602, 54518 Vandœuvre-lès-Nancy, France

<sup>b</sup> Tecnológico Nacional de México/Instituto Tecnológico de Veracruz: Calz. M.A. de Quevedo 2779, Veracruz, Ver, Mexico

## HIGHLIGHTS

- Monitoring of mAb glycosylation during CHO cell processes by *in situ* NIR probes.
- PLS regressions are limited for providing spectrum-based glycosylation information.
- Consideration of the process nature in regression is critical for accurate monitoring.
- The novel use of LWR allowed to monitor total mAb concentration accurately.
- Only SVR allowed mAb glycoform concentration monitoring.

## ARTICLE INFO

## Keywords:

Antibody real-time glycosylation monitoring  
Support vector regression  
Locally weighted regression  
Process Analytical Technology  
Quality by Design

## ABSTRACT

Since monoclonal antibodies (mAb) are sensitive to the manufacturing process, several mAb variants can be the result of a single batch production. The most critical source of heterogeneity is glycosylation which has a profound impact on safety and efficacy of the final product. Implementation of monitoring and control of the process using the Quality by Design (QbD) approach may help to ensure mAb specifications, although its implementation is limited by the availability of real-time specific measurements. All current approaches to elucidate mAb glycoforms require sampling and labour-intensive efforts. Thus, glycosylation analysis is often performed with the objective of detecting quality defects at the end of the culture process. In this work, the capability of Near Infrared spectroscopy and chemometric treatment to accurately monitor mAb glycosylation during CHO cells cultures using *in situ* probes is shown for the first time. Real-time monitoring of glycosylation, in terms of high mannose isoforms, fucosylated, sialylated and galactosylated isoforms as well as non-glycosylated mAb, has been successfully performed by the novel use of Locally Weighted Regression (LWR) and Support Vector Regression (SVR). These encouraging results open the way for the implementation of control systems on the impact of cell culture operating parameters on mAb heterogeneity, particularly glycosylation, during CHO cell culture processes through the QbD approach.

## 1. Introduction

Monoclonal antibodies (mAb) produced in animal cell culture processes represent a success in terms of clinical benefit for patients and

revenue generated by biopharmaceutical industries. Such molecules are quite sensitive to changes in manufacturing processes and thus several mAb variants could be produced within a single batch due to post-translational modifications. Glycosylation is the main source of mAb

\* Corresponding authors.

E-mail addresses: [daniel.zavala-ortiz@univ-lorraine.fr](mailto:daniel.zavala-ortiz@univ-lorraine.fr), [daniel.zavala@itver.edu.mx](mailto:daniel.zavala@itver.edu.mx) (D.A. Zavala-Ortiz), [bruno.ebel@univ-lorraine.fr](mailto:bruno.ebel@univ-lorraine.fr) (B. Ebel), [mengyao.li@univ-lorraine.fr](mailto:mengyao.li@univ-lorraine.fr) (M.-Y. Li), [dulce.bd@veracruz.tecnm.mx](mailto:dulce.bd@veracruz.tecnm.mx) (D.M. Barradas-Dermitz), [patricia.hj@veracruz.tecnm.mx](mailto:patricia.hj@veracruz.tecnm.mx) (P.M. Hayward-Jones), [gaguilar@itver.edu.mx](mailto:gaguilar@itver.edu.mx), [maria.au@veracruz.tecnm.mx](mailto:maria.au@veracruz.tecnm.mx) (M.G. Aguilar-Uscanga), [annie.marc@univ-lorraine.fr](mailto:annie.marc@univ-lorraine.fr) (A. Marc), [emmanuel.guedon@univ-lorraine.fr](mailto:emmanuel.guedon@univ-lorraine.fr) (E. Guedon).

<https://doi.org/10.1016/j.bej.2019.107457>

Received 20 August 2019; Received in revised form 16 October 2019; Accepted 22 November 2019

Available online 29 November 2019

1369-703X/ © 2019 Elsevier B.V. All rights reserved.

variability which can strongly impact mAb clinical properties. Therefore, the control of glycosylation specific profiles of mAb during the process is critical for therapeutic efficacy and patient safety. For this purpose, regulatory agencies proposed the Process Analytical Technology (PAT) strategy to control pharmaceutical manufacturing processes through the continuous adjustment of Critical Process Parameters (CPP) which affect Critical Quality Attributes (CQA) of the product. Accordingly, continuous monitoring of CQA, such as mAb glycosylation, is required to establish advanced retro-control systems to guarantee mAb specifications [1].

However, continuous monitoring of mAb glycosylation is challenging since it requires analyses at a relatively high cost. Complete mAb glycosylation analysis must include identification of sugar chains attached to glycosylation sites (macro-heterogeneity) and of the sugars moieties within the glycan chain (micro-heterogeneity). Indeed, mAb macro- and micro-heterogeneity analyses involve several steps, such as enzymatic digestion, labelling, derivatization and separation, followed by structural analysis, usually using mass spectrometry (MS) [2]. The complexity of such analyses implies significant delays, 5 h to 2 days, mainly for the acquisition of process information, thus mAb quality analysis is usually performed at the end of the cell culture process [3].

In recent years, intensive efforts have been made to establish PAT as the mean to monitor mAb glycosylation during cell cultures, particularly using automatic at-line or on-line traditional biochemistry approaches with the aim of reducing analysis times and sample volumes. Such approaches allow the presence of mAb glycoforms to be known in a question of hours [4–7]. In the last few decades, vibrational spectroscopy, in combination with multivariate analysis, has been proven to be a promising tool [8], particularly for cell culture monitoring [9]. However, it is mainly restricted to some usual cell substrates and/or by-products such as glucose, lactate or glutamine [10–15]. More recently, a study showed the possibility of monitoring mAb concentration by using either Raman or NIR spectroscopy in real-time during CHO cell cultures [16]. Although Raman spectroscopy led to a slightly better estimation for mAb concentration, NIR spectroscopy showed a higher signal-to-noise ratio, though in more complex spectra. The inferior capacity of the NIR model was thus mainly attributed to the lack of linear PLS regression for handling complex NIR spectra, likely containing information in non-linear ways [16]. However, as far as can be ascertained, glycosylation micro-heterogeneity monitoring has not been addressed. In this study, we showed that *in situ* NIR spectroscopy can be applied for in-line monitoring of mAb glycosylation micro-heterogeneity.

As alterations in mAb glycoform patterns may result in strong changes in clinical profiles, manufacturers must guarantee glycosylation specifications to ensure reproducible and consistent clinical performance. The occurrence of several factors that may compromise conservation of clinical profiles is not uncommon in the pharmaceutical industry [17], and so monitoring and control systems are required to ensure mAb properties. Conservation of such properties is due to proper combination of glycoforms with different sugar moieties within the glycan chain. Indeed, it has been widely reported that the presence of fucose, galactose and sialic acid strongly affects antibody dependent cellular cytotoxicity (ADCC), complement dependent cellular cytotoxicity (CDC) and immune modulation of the mAb, respectively [18]. On the other hand, high mannose glycoforms are reported to reduce serum half-life [19]. In this context, producing mAb glycoforms with resulting clinical effects similar to those of the reference mAb is critical for batch approval [17]. In this study, it was shown that NIR spectra can exhibit an estimated correlation to non-glycosylated mAb and total mAb concentration (mAb macro-heterogeneity), as well as to glycoforms containing fucose, galactose and high-mannose structures, including sialic acid within the glycan chains (mAb micro-heterogeneity). Furthermore, this approach could have an immediate application using a NIR spectrophotometer in at-line or off-line modes, which could provide mAb glycosylation information in question of minutes.

## 2. Theory

Spectra, particularly from NIR in-line analysers, are complex since both physical and chemical information is contained, usually in a highly collinearity way. Thus, multivariate calibration particularly for regression, is needed for correlating complex spectra to desired quality attributes or analyte concentration. Though many regression methods are available for building calibration models, as far as it can be ascertained, only Partial Least Squares Regression (PLSR) has been addressed in cell culture monitoring. In this work we analysed the adequateness of PLSR for cell culture monitoring and explored the performance of other regression methods such as Support Vector Regression (SVR) and Locally Weighted Regression (LWR). The intuitive concept of models as well as their characteristics for regression in cell cultures are discussed.

### 2.1. Partial Least Squares Regression (PLSR)

The PLSR method is based in a reduction variable process in order to treat collinearity. Firstly the spectra (X) and the concentration (Y) matrix is decomposed as [20]:

$$X = TP^T + R_1 \quad (1)$$

$$Y = UQ^T + R_2 \quad (2)$$

where X and Y are spectra and concentration matrices respectively, T and U are the pseudo-scores matrices, P and Q are the pseudo-loadings matrices and  $R_1$  and  $R_2$  are the residuals matrices. Matrix decomposition of X and Y matrices is not independent, thus an internal relationship between the scores of X and Y is generated accordingly:

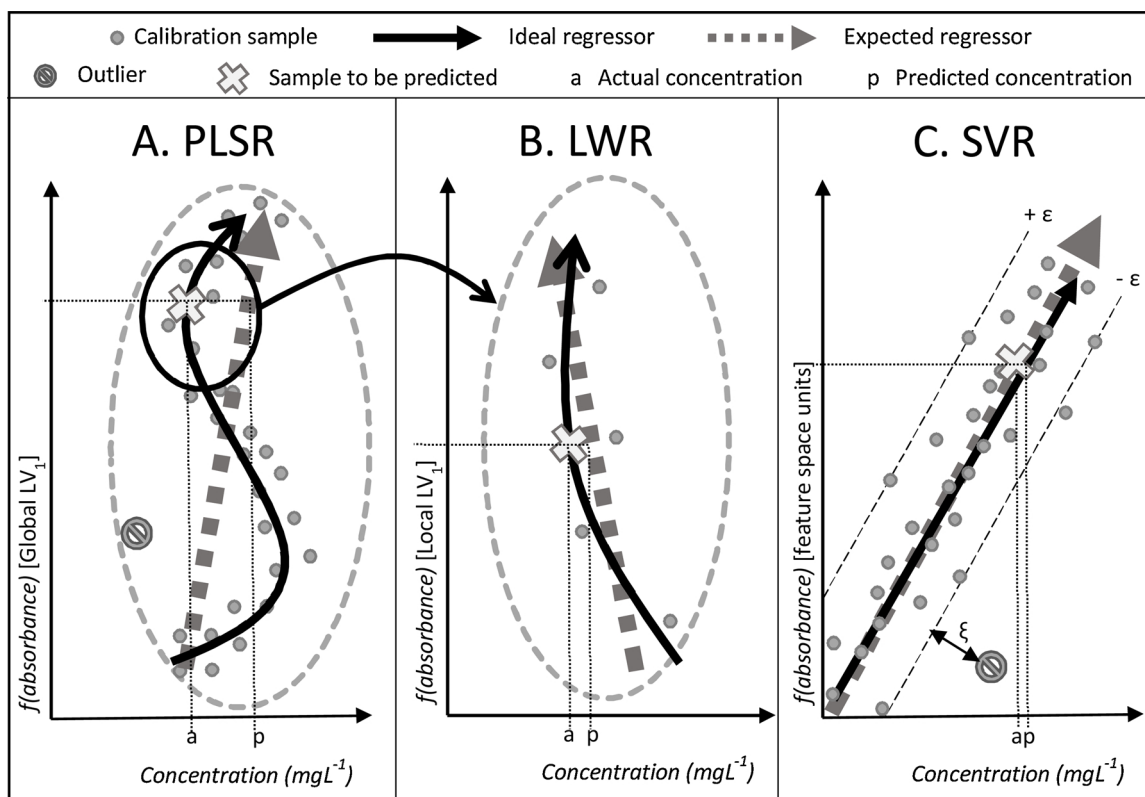
$$U = BT \quad (3)$$

where U is the pseudo-scores of Y to be calculated, T the pseudo-scores of X and B the regressor matrix. PLS works with the constraint that these components explain as much as possible of the covariance between X and Y. Once the regressor matrix has been determined, calculation of y-concentration value from problem sample may be calculated as:

$$Y = T^*BQ^T + R_1 \quad (4)$$

Where  $T^*$  is the pseudo-scores matrix of the problem sample, B the regressor matrix,  $Q^T$  the pseudo-loading matrix of the model and  $R_1$  the residual matrix. Regression based on PLSR offers a relatively simple frame for analysing the relationship between spectral response and prediction by the model. However, it is particularly sensitive to scattering effects and may not properly handle information contained in a non-linear way. Therefore, for building quantitative calibration models using NIR spectroscopy, a common assumption has been focusing on chemical information and limiting the contribution of physical data contained in spectra by the use of spectral pre-treatment [21]. However, manipulation of spectra for limiting scattering effects may also disturb chemical information [22] and thus a compromise must be adopted.

PLSR is a variable space-based regression method which calculates the relationship between each of the variables (absorptions at different wavelengths) and compound concentrations. Therefore, such a relationship should be relatively constant during the whole culture process in order to maintain accurate estimations. Though it is an obvious assumption, caution must be taken considering the strong physical and chemical variations of culture media during cell culture progression. Deviations from this assumption are usually observed as non-linear effects (Fig. 1-A) and may limit the predictive capability of models. In such cases, the regression equation leads to a linear trajectory of predicted values (grey dotted arrow in Fig. 1-A), limiting prediction on non-linear sections (cross in Fig. 1-A). As far as can be ascertained, this fact has not been addressed yet in cell culture monitoring. Therefore, the novel use of space-based regression methods which firstly focus on



**Fig. 1.** Intuitive concept of regression methods. PLSR (A), LWR (B), SVR (C). PLSR creates a chemometric space in term of latent variables (LV) and then a linear regression is performed for estimating concentration. LWR enhances the fitting of non-linearities by performing local regression by weighted PLSR. In contrast to PLSR which generates the regressor by only minimising the error, SVR maps spectra into a higher dimensional feature space as for fitting the error within a particular threshold ( $\pm \epsilon$ ) and then the regressor is generated.

affinity and dissimilarity between samples (culture progression information), and secondly on the relationship between variables (absorptions at different wavelengths), has been evaluated.

## 2.2. Locally Weighted Regression (LWR)

LWR is mainly addressed for modelling complex relationships for which no theoretical model may exist. In contrast to PLSR which generates a regression function considering all calibration points, LWR firstly compares the sample to be predicted with samples within the calibration set. Then only those calibration samples similar to the sample to be predicted are used (the local area) (black circle in Fig. 1-A). Then each point of the local area is weighted according to its distance from the sample to predict: close points are given more importance or weight, far points are given less weight; then a regression function of the independent variables is generated employing the weights and in the local area [23] (Fig. 1-B). Generation of accurate models then requires adjustment of key parameters such as similarity between samples, definition of the local area and the weights, and the nature of the regression itself.

Once spectra have been mapped into a chemometric space (principal component space, latent variable space, among others), the local area is determined by a distance function and specified limit. Since distance as a delimiting criterion may be inappropriate when lacking vast calibration samples in a wide calibration space, several authors have employed distance criteria in terms of near calibration samples [24,25]. Then local calibration samples are weighted according to a weight function, such as the tricubic function (5). Then the weight for a calibration sample is calculated as:

$$W(u) = \begin{cases} (1 - u^3)^3 & \text{if } u \leq 1 \\ 0 & \text{if } u > 1 \end{cases} \quad (5)$$

$$w_i(x_j) = W\left(\frac{\delta(x_j, x_i)}{d(x_j)}\right) \quad (6)$$

where

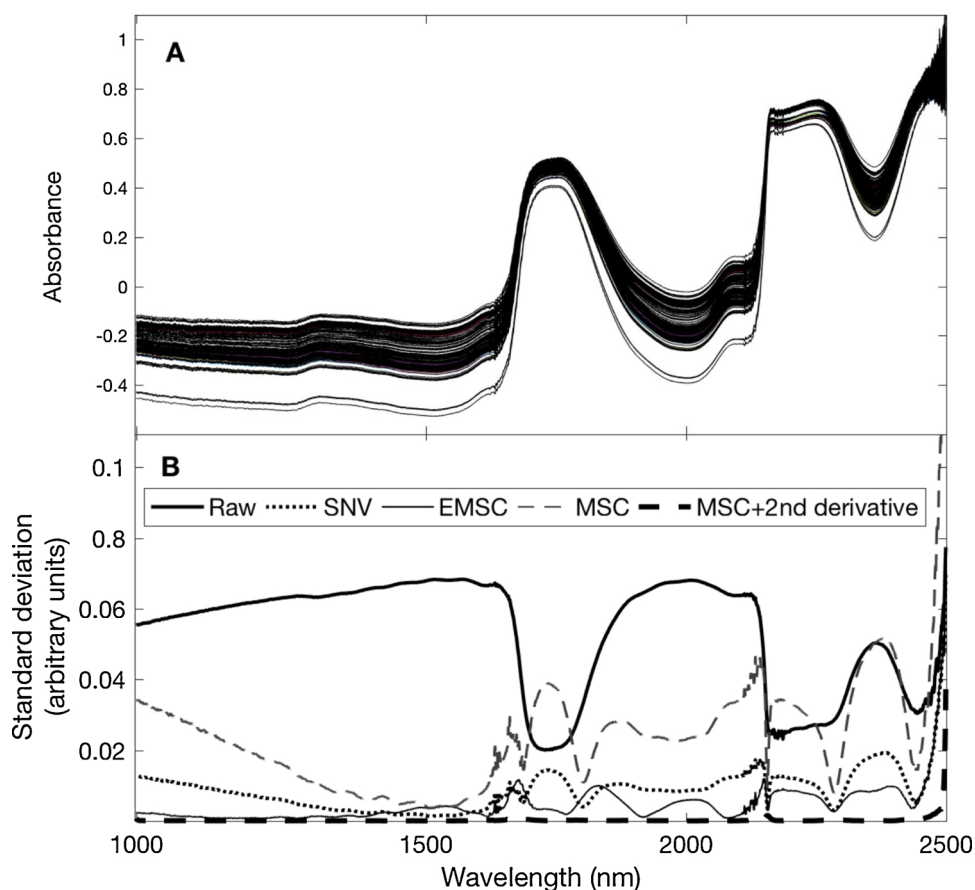
$(x_j, x_i)$ : Distance between prediction sample  $j$  and calibration sample  $i$

$d(x_i)$ : Maximum distance involved in each regression

As could be noted, the weights will be large (close to 1) for  $x_i$  close to  $x_j$ , and small (close to 0) for  $x_i$  far from  $x_j$ . Once the region and weights have been determined, regression function in the local region is generated generally using weighted PLSR. Finally, the concentration value for  $x_j$  is calculated using the local weighted PLSR regressor (grey dotted arrow in Fig. 1-B). In contrast to global PLSR that treats all the regression surface at the same time, as either linear or non-linear, LWR models non-linear regions without compromising linear regions. This approach is particularly adequate for animal cell culture processes in which linear and non-linear behaviour may arise differently during different phases of cell cultures. Moreover, the use of similar samples in the local area could lead to better spectra pre-treatment and thus limiting the loss of information by attenuation of scattering effects in spectra. Drawbacks of LWR are the need of dense calibration samples, vulnerability to outliers [26] and the lack of a mechanistic model where fitted parameters specify particular physical or chemical properties of the cell culture. This is of great concern since regulatory agencies demand that NIRS signals be directly attributed to analytes or be an indirect measurement correlated with light scattering effects [27,28]. Then submission of monitoring procedures would eventually require efforts considering all possible combinations of local regressions.

## 2.3. Support Vector Regression (SVR)

A relatively novel alternative for non-linear modelling of NIR



**Fig. 2.** Nature of spectra for calibration. Process instrument raw spectra for calibration (A), effect of common spectral pre-treatments on spectra variability and scattering effects, MSC: Multiplicative Scatter Correction, SNV: Standard Normal Variate, EMSC: Extended Multiplicative Scatter Correction (B).

spectra is SVR [29]. The main difference of SVR from other typical regression methods is that its objective is not merely to reduce the fitting error but to fit the error within a particular threshold ( $\pm \varepsilon$ ). Then the goal of SVR is to generate a regression function, or hyper plane, that has a maximum number of calibration samples at most an  $\varepsilon$  deviation from an actual concentration ( $y_i$ ), and at the same time keeping the function as flat as possible [30]. For instance, the hyper plane is considered as:

$$f(x) = (wx) + b \quad (7)$$

with  $w \in X$ ,  $b \in \mathbb{R}$ , and  $x$  being a variable related to spectra

Flatness is then assured by minimisation of  $w$ , for example minimising the norm as a convex optimization problem:

$$\text{minimise } \frac{1}{2} |w|^2 \quad (8)$$

subject to

$$y_i - (wx) - b \leq \varepsilon \text{ and } (wx) + b - y_i \leq \varepsilon \quad (9)$$

However, it may not be the case that  $f(x)$ , which approximates all pairs  $(x_i, y_i)$  with  $\varepsilon$  precision, actually exists. Then a soft margin of slack variables  $\xi_i, \xi_i^*$ , is introduced for coping with unfeasible constraints of optimization (8) as stated by Vapnik [31]:

$$\text{minimise } \frac{1}{2} |w|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (9)$$

subject to

$$y_i - (wx) - b \leq \varepsilon + \xi_i, \quad (wx) + b - y_i \leq \varepsilon + \xi_i^* \text{ and } \xi_i, \xi_i^* \geq 0 \quad (10)$$

The constant  $C$  determines the compensation between the flatness of

$f(x)$  and the amount up to which deviations larger than  $\varepsilon$  are tolerated. This general procedure is depicted in Fig. 1-C. As could be observed, SVR is then less vulnerable to outliers since it could properly generalise and leave the outliers in the soft margin (forbidden symbol in Fig. 1-C). In complex multivariate data optimization, (9) can be solved more easily in its dual formulation, which provides the possibility for extending the procedure to non-linear functions. This could be achieved by mapping the  $x_i$  patterns into some feature space  $F$  [32]:

$$\varphi: x \rightarrow F \quad (11)$$

Then standard SVR procedure is applied. Mapping into a higher, linear or non-linear, dimensional space, may require exacerbated computational power, thus the majority of SVR uses implicit mapping by kernels. The most common are linear, polynomial and Gaussian radial basis function (RBF) kernels. The nature of the calibration set must be considered for properly selecting the kernel [33]. The linear kernel is useful in large sparse data vectors with linear regularization, the polynomial may fit some soft non-linearity and RBF are general-purpose that are generally applied in strong non-linear regularization or in the absence of prior knowledge [33]. This approach could be used for generalizing difficult to fit data in complex systems. As well as for LWR, monitoring procedure submission could be likely cumbersome for relating SVR parameters to specific chemical or physical properties of the cell culture, particularly in strong non-linear processes mapped into high dimensional feature space.

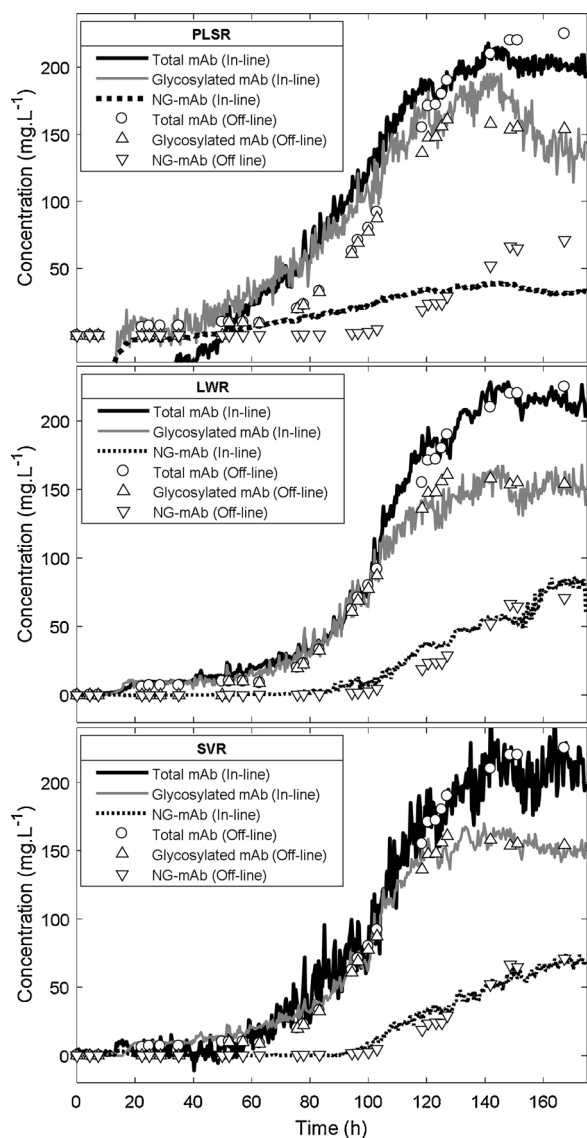


Fig. 3. Comparison of PLSR, LWR and SVR models to monitor mAb macro-heterogeneity glycoforms concentration by *in situ* NIR spectroscopy, during batch CHO cell culture. Real time estimation by models (lines) and off-line measured values (symbols) for the groups of (—,○) total mAb, (—,△) glycosylated mAb and (.....,▽) NG-mAb.

### 3. Materials and methods

#### 3.1. Cell cultures for NIR spectra acquisition

The bioreactor data set was designed with routinary monitoring for batch culture in mind. Several cultures of CHO cells were performed in 2 L bench-top bioreactors (Pierre Guérin, France) with a 1.5 L working volume: three batch cultures, two feed-harvest cultures with medium renewal and one batch culture with glucose spiking. The three batch cultures were intended to observe inter-batch heterogeneity as well as in-line and off-line expected routinary responses. Feed-harvest cultures were used for increasing the variance of mAb glycoforms within the calibration process, which could enhance model prediction capability. These were started after a first phase in batch mode, then 2/3 of cell culture was withdrawn and replaced by fresh culture medium. This procedure was repeated 2 and 4 times for these 2 feed-harvest cultures respectively. As relative abundance of particular mAb glycoforms is partially a function of cell culture progression, the use of feed-harvest cultures favoured not only samples with mAb at the beginning of the

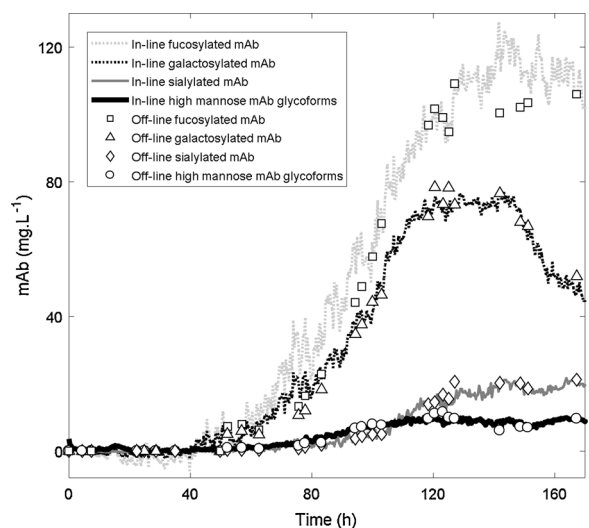


Fig. 4. Performance of SVR models to monitor mAb micro-heterogeneity glycoforms concentration by *in situ* NIR spectroscopy, during batch CHO cell culture. Real time estimation by models (lines) and off-line measured values (symbols) for the groups of (.....,□) fucosylated mAb, (.....,△) galactosylated mAb, (.....,◇) sialylated mAb and (—,○) high mannose mAb.

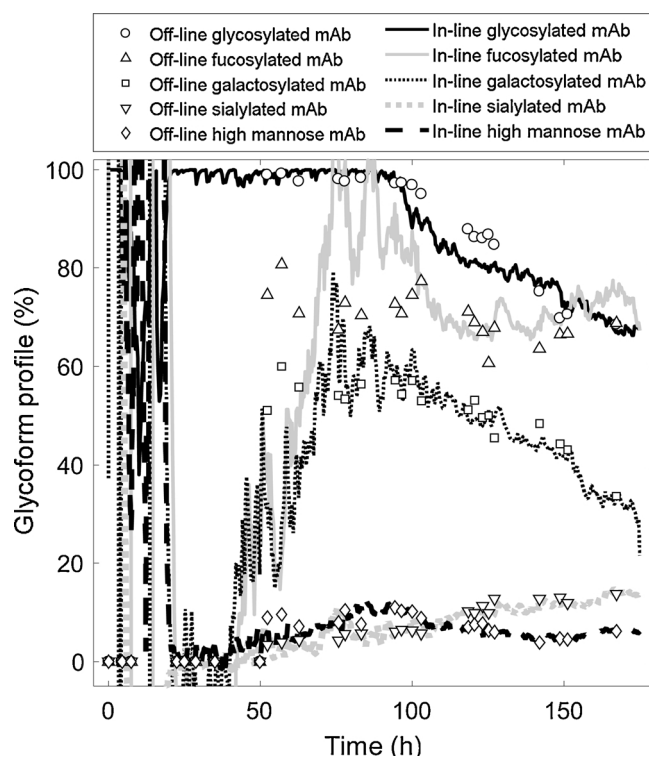


Fig. 5. Example of model performances for automatic in-line monitoring of mAb glycosylation profiles. The mAb macro-heterogeneity profile was calculated as the concentration ratio between glycosylated mAb (estimated by SVR model) and total mAb (estimated by LWR model). The mAb micro-heterogeneity profiles were calculated as the ratio of fucosylated, galactosylated, sialylated and high mannose glycoforms in relation to glycosylated mAb concentration (all estimated by SVR models). Real time estimation (lines) and off-line values (symbols) for the groups of (—,○) glycosylated mAb, (—,△) fucosylated mAb, (.....,□) galactosylated mAb, (.....,▽) sialylated mAb and (—,◇) high mannose mAb.

cultures, but also cell cultures with a wider variability of mAb glycoforms. Batch culture with glucose spiking was used for increasing mAb concentrations so that these values during routinely batch culture

**Table 1**  
Spectral pre-treatment used for models.

Compound	Regression method	Pre-treatment
Total mAb	PLSR	EMSC + SNV
	LWR	SNV
	SVR	EMSC
NG-mAb	PLSR	EMSC
	LWR	Detrend + MSC
	SVR	MSC
Glycosylated-mAb	PLSR	EMSC
	LWR	EMSC
	SVR	MSC
F-glycoforms	SVR	MSC
G-glycoforms	SVR	MSC
S-glycoforms	SVR	MSC
HM-glycoforms	SVR	MSC

NG-mAb: Non-glycosylated mAb; G-mAb: Glycosylated mAb; F-glycoforms: Glycoforms containing Fucose; G-glycoforms: Glycoforms containing galactose; S-glycoforms: Glycoforms containing sialic acid; HM-glycoforms: Glycoforms of high mannose structures.

EMSC: Extended Multiple Scatter Correction, SNV: Standard Normal Variate, MSC: Multiple Scatter Correction.

Auto scale was always applied as last pre-treatment step.

**Table 2**  
Performance of models during calibration.

Model structure	Rang(mg.L <sup>-1</sup> )	RMSECV (mg.L <sup>-1</sup> ) (%)	R <sup>2</sup>	Compound
PLSR: 5 LV	0 - 380	34.1 9	0.85	Total mAb
LWR: 4 LV, 21 LP		31.1 8	0.90	
SVR: 120 SV		32.4 8	0.87	
PLSR: 3 LV	0 - 98	19.0 19	0.38	NG-mAb
LWR: 12 LV, 15 LP		15.0 15	0.61	
SVR: 148 SV		10.4 11	0.70	
PLSR: 3 LV	0-330	47.1 14	0.70	Glycosylated-mAb
LWR: 3 LV, 15 LP		37.5 11	0.81	
SVR: 158 SV		33.1 10	0.83	
PLSR: 4 LV	0 - 262	37.6 14	0.68	F-glycoforms
LWR: 24 LP, 3 LV		28.3 11	0.82	
SVR: 167 SV		27.9 11	0.83	
PLSR: 5 LV	0 - 147	17.6 12	0.71	G-glycoforms
LWR: 20 LP, 5 LV		13.7 9	0.83	
SVR: 166 SV		13.3 9	0.85	
PLSR: 4 LV	0 - 19	2.4 13	0.70	S-glycoforms
LWR: 12 LP, 4 LV		2.2 12	0.74	
SVR: 148 SV		1.9 10	0.81	
PLSR: 5 LV	0 - 47	6.8 14	0.62	HM-glycoforms
LWR: 15 LP, 3 LV		5.4 12	0.77	
SVR: 166 SV		4.6 10	0.83	

NG-mAb: Non-glycosylated mAb; G-mAb: Glycosylated mAb; F-glycoforms: Glycoforms containing Fucose; G-glycoforms: Glycoforms containing galactose; S-glycoforms: Glycoforms containing sialic acid; HM-glycoforms: Glycoforms of high mannose structures. LV: Number of Latent Variable used for the PLSR and LWR models, LP: Number of local points used for the LWR models, SV: Number of Support Vectors used by the SVR models.

**Table 3**  
Glycoforms considered for NIR mAb micro-heterogeneity calibration models.

NIR models	Fucosylated	Galatosylated	Sialylated	High Mannosylated
Glycoforms	G0F, G1F, G2F, G0F-N, G1F-N, G0FN, G1FN, G2FN, G1FS, G2FS, G2FS2, G1FS-N	G1F, G2F, G1F-N, G1, G2, G1FN, G2FN, G2N, G1FS, G2FS, G2FS2, G1S, G2S, G1FS-N	G1FS, G2FS, G2FS2, G1S, G2S, G1FS-N	Man5, Man6, Man7, Man8

Oligosaccharide structures attached to the conserved site (Asn 297) in each heavy chain of the mAbs.

F a fucose moiety in the sugar chain; G a galactose moiety in the sugar chain: 0 no galactose, 1 a galactose in only one branch, 2 if both branches galactosylated; S a sialic acid in the sugar chain: no number if a S in only one branch, 2 if both branches contain a S; N N-acetylhexosamine in the sugar chain; Man mannose moieties in the sugar chain, number indicating the number of mannose moieties in the sugar chain.

monitoring relied preferably within an appropriate concentration range.

The culture medium was a protein-free medium mixture consisting of a 1:1 vol ratio of PF-CHO (HyClone) and CD-CHO (Fisher Scientific) supplemented with 4 mM L-glutamine (Sigma Aldrich) and 0.1 % pluronic F-68 (Sigma Aldrich). The genetically modified DG44 CHO cell line was used (human anti-Rhesus D mAb-producing CHO M250-9), kindly provided by Bioprocessing Technology Institute (Singapore). Dissolved oxygen (DO) was controlled at 50 % air saturation and agitation rate was fixed at 90 rpm throughout the culture. Temperature was maintained at 37 °C and pH was set and controlled at 7.2 using 0.5 M sodium hydroxide and CO<sub>2</sub>.

*In-situ* (or in-line, invasive) spectral scanning of bioprocess culture media was carried out with a NIR transfectance probe with 1 mm pathlength (Precision Sensing Devices, MA). The autoclavable probe was connected to an Antaris II spectrometer (Thermo Scientific, USA). Each NIR spectrum corresponded to an average of 128 scans from 1,000–2,500 nm.

### 3.2. Off-line analyses

Off-line concentration of total mAb was determined using an enzymatic kit (Roche Life Science) with an automatic spectrophotometer (Thermo Scientific GALLERY) against external standards. The nature and concentration of mAb heterogeneity in the form of glycoforms was elucidated by HPLC/UHPLC-mass spectroscopy analysis, as previously described [15]. Off-line concentration values for calibration included and exceeded those expected during routine monitoring of batch cultures (0–240 mg.L<sup>-1</sup> and 0–75 mg.L<sup>-1</sup> for total mAb and NG-mAb respectively) and also the variability of mAb glycoforms. Off-line total mAb concentration range used for calibration was 0–380 mg.L<sup>-1</sup>, off-line NG-mAb range was 0–98 mg.L<sup>-1</sup>. Analysis of mAb glycoform relative abundance profiles revealed a significant difference between exponential and stationary-death phase of cultures (One way ANOVA,  $p \leq 0.05$ ), particularly for NG-mAb, G0F, G1F, G2F and Man5 mAb glycoforms (data not shown). Moreover, enhancement of prediction capacity is expected since the use of feed-harvest cultures increased mAb glycoform variability during the calibration process.

### 3.3. Development and analysis of calibration models

Firstly, in order to generate the calibration methods for mAb glycoforms, special attention was given to spectrum pre-processing according to Huang [22]. The presence of additive, multiplicative and wavelength-dependent effects due to scattering was evaluated within the calibration set spectra. The most common techniques to eliminate undesired spectral variations caused by light scattering (Multiplicative Scatter Correction-MSC, Probabilistic Quotient Normalization-PQN, Standard Normal Variate-SNV, Extended Multiplicative Scatter Correction-EMSC, derivatives) were evaluated. The standard deviation for each wavelength was used to elucidate the effect of scattering on calibration spectra; this data was considered for final spectral pre-treatment. The calibration set comprised 168 spectra collected from six bioreactor cultures. PLSR, LWR and SVR models and statistical analysis

were performed in MATLAB® (Statistics and Machine Learning Toolbox™, MATLAB R2016a, The MathWorks, Inc., Natick, Massachusetts, United States) using chemometric software (PLS\_Toolbox® 8.2.1, Eigenvector Research, Inc., Manson, WA, United States). Model performance was evaluated for accuracy by Root Mean Square Error of cross-validation (RMSECV) and square correlation coefficients ( $R^2$ ). A low value of RMSECV is related to enhanced accuracy, while a high value  $R^2$  value indicates that the model properly handles spectrum variability to perform concentration estimation.

Firstly, PLSR models were performed using a venetian blinds cross-validation. Determination of latent variable (LV) number was based on the goodness of estimation (Q2Y): the minimum number of LVs was obtained when Q2Y ceased to improve. LWR was applied to fit global non-linear relationships by local linear regressions using PLSR and the classic cubic weight equation. Determination of local areas in term of local points, and LV number, was performed by optimization of these parameters with RMSECV as the response variable. For SVR models, an epsilon-support vector regression using a Gaussian radial basis function kernel was used. SVR models were optimized using a random subset cross-validation approach with maximal error values corresponding to deviations up to 10 % from actual values.

### 3.4. *In situ* monitoring of mAb glycoforms

The focus of the work was primarily to analyse predicted or estimated kinetic profiles using batch culture. This provides a frame containing different physiological cell states within lag, exponential, stationary, and death phases of batch culture which dynamically impact the nature of mAb glycoform profiles. This strategy may be useful to infer model performance in different matrix compositions, which may help in future work.

Once calibration models were optimized, they were used independently to perform *in situ* monitoring of mAb glycosylation during a CHO cell culture. Characterization of mAb was carried out in terms of macro-heterogeneity (total mAb concentration, glycosylated mAb and non-glycosylated mAb) and micro-heterogeneity (high mannose glycoforms and glycoforms containing any fucose, sialic acid and galactose moiety). A NIR analyser was programmed for performing automatic *in situ* scanning of culture medium every 20 min. Batch culture monitoring produced 500 spectra from which only 27 were used for calibration. Thus approximately 95 % NIR data was not used to establish the models and may further depict the prediction performances of models. For the evaluation of monitoring, models mimic real time monitoring of the mAb producing cell culture process. Spectra captured every 20 min were then used as inputs. Then calibration models returned mAb concentration values that were used to real time generate the kinetic profiles of mAb macro- and micro-heterogeneity (Figs. 3–5). As global therapeutic effects of mAb is mainly a function of the micro-heterogeneity profile of the lot, real-time glycosylation data from best models were used to real-time monitor the global glycosylation profile of the produced lot. Firstly, the macro-heterogeneity profile was determined as the relationship of NG-mAb concentration estimated by the SVR model, with total mAb concentration estimated by the LWR model. Secondly, the micro-heterogeneity profile was determined as the relationship of fucosylated, galactosylated, sialylated and high mannose glycoforms with glycosylated mAb concentration, using SVR models.

## 4. Results and discussion

### 4.1. Development and analysis of NIR models based on PLSR, LWR and SVR methods

Spectra for calibration were evaluated for scattering effects (Fig. 2-A). The most common techniques to eliminate undesired spectral variations caused by light scattering (MSC, SNV, EMSC, derivatives) were evaluated. General analysis of spectra revealed some scattering effects

such as additive effect (baseline shift), multiplicative effect (offset of spectra) and a likely wavelength-dependent effect from approximately 1000–1500 nm (Fig. 2-B). These effects were particularly observed in spectra of stationary and cell death phases where maximum mAb concentration was achieved (data not shown). Only EMSC was effective for limiting the likely wavelength-dependent effect, also the use of derivatives with any normalization pre-processing, particularly 2nd order derivative with MSC (Fig. 2-B). This analysis was firstly used for selecting spectra pre-treatments for models. After spectral analysis, random trials of promising pre-treatments and their combinations were assessed for reducing RMSECV of models.

In general terms EMSC led to PLSR models with lower RMSECV, likely due to a proper compromise between reduction of spectra variability and scattering effects, particularly multiplicative and wavelength-dependent effects. The use of derivatives after normalization, limited the predictive power of PLSR models (data not shown), likely due to a strong reduction on spectra variability (Fig. 2-B) and thus reduction on chemical information. On the other hand, LWR and SVR showed greater management of scattering effects, particularly wavelength-dependent effect since only MSC was sufficient for reduction of RMSECV values in almost all models. Final spectral pre-treatments used for calibration are shown in Table 1. Construction of mAb glycosylation models was performed using different regression methods (PLSR, LWR and SVR) as reported in Materials and Methods. Performances of models during calibration are summarized in Table 2. PLSR models led to poor estimation capability, even for total mAb concentration. PLSR is a variable space-based regression method which calculates the relationship between each of the variables (absorptions at different wavelengths) and compound concentrations. Such a relationship should be relatively constant during the whole culture process in order to maintain accurate estimations, including the scattering nature of the matrix. Perhaps the most evident deviation from this assumption is the fact that scattering effects occurred and impacted spectra in different ways, depending on the increase in scattering compounds (cells, cell debris, among others) according to batch culture progression. A plausible reason for poor performance is the limited capacity of PLSR for handling multiplicative and wavelength-dependent effects [34], likely caused by scattering compounds. Handling spectra with such different scattering natures with the same spectral pre-processing as commonly done in PLSR, would not only lead to correcting response in a narrow frame but also masking chemical information in the uncorrected frame.

As a result, LWR and SVR were, for the first time, evaluated for cell culture monitoring. LWR and SVR are sample space-based regression methods which firstly focus on affinity and dissimilarity between samples (batch progression information), and secondly on the relationship between variables (absorptions at different wavelengths) and compound concentrations. LWR uses only similar samples in the PLS space to perform local regression using weighted PLSR, while SVR consists of a number of support vectors corresponding to samples from the calibration set and non-linear model coefficients defining the relationships between spectra and compound concentrations. As LWR and SVR use only similar samples with a similar matrix nature, including similar scattering effects, it is likely that non-desirable effects of spectral pre-processing are limited.

Results showed that, in contrast to PLSR performance, SVR and LWR were superior for estimating the concentration of all glycoforms (Table 2). This can be explained by the fact that SVR and LWR not only consider the relationship between spectra and compound concentrations, but also cell culture progression in terms of cell density, viability and metabolite concentrations. With the exception of total mAb which was better estimated by LWR, SVR was likely the best option for all mAb glycoforms, particularly glycosylated mAb.

### 4.2. Real time monitoring of mAb glycosylation

Firstly, mAb glycosylation macro-heterogeneity monitoring was



addressed as shown in Fig. 3. As expected, evaluation of PLSR models revealed a limited capacity for monitoring mAb macro-glycoforms, particularly non-glycosylated mAb, due to non-linear relationships between spectra and non-glycosylated mAb concentration (data not shown). In fact, non-linear relationships are likely the result of physical (scattering, mass and heat dynamics) and chemical (chemical composition changes) phenomena that strongly change the interaction of NIR radiation with mAb during progression of batch cell cultures. LWR breaks global nonlinearity by performing several local regressions using only similar samples. In this context, LWR was successful in monitoring total and glycosylated mAb concentration (Fig. 3) that display some non-linearity mainly associated to physical phenomena. However, for non-glycosylated mAb, only trends were observed. Limited capacity to estimate NG-mAb concentration by the LWR model is explained by the fact that an inherent nonlinear relationship between spectra and concentration existed (data not shown), which cannot be properly modelled by the local linear regressions. As shown in Fig. 3, the novel use of LWR and SVR as enhanced regression methods, allowed the proper monitoring of total mAb and NG-mAb respectively. These results demonstrated the capability to monitor mAb glycosylation macro-heterogeneity in real-time, using *in situ* NIR spectroscopy. As more accurate and stable estimations of glycosylated mAb concentration were achieved using SVR, calibration for glycosylated glycoforms was addressed using the SVR approach.

As for mAb micro-heterogeneity, among a total of 25 potential glycosylated mAb glycoforms reported for mAb produced in CHO cell cultures [2], only some glycoforms were detected off-line (data not shown). Thus, mAb micro-heterogeneity models were generated based on the detected glycoforms that contained particular sugar moieties conferring clinical properties. Detected glycoforms were classified into 4 groups for mAb micro-heterogeneity model development, corresponding to high mannose, fucosylated, sialylated and galactosylated isoforms (Table 3). As a glycosylated chain may contain different sugar moieties, it is possible that one particular glycoform be considered for two or more calibration models. SVR was capable of properly extracting mAb glycosylation information from NIR spectra, which allowed mAb micro-heterogeneity monitoring as shown in Fig. 4. Even sialylated and high mannose glycoforms whose concentrations were low ( $< 15 \text{ mg.L}^{-1}$ ), were specifically detected. These results demonstrated the capability of *in situ* NIR spectroscopy to quantitatively monitor mAb micro-heterogeneity.

According to the QbD initiative, real time monitoring should finally be used for performing advanced retro-control. Therefore, concentration values of mAb glycoforms must also be monitored using a reference frame as for setting target values for glycoforms ratios, which are related to mAb clinical effects. Monitoring of a process under this approach is a more challenging task since calculation of ratios could increase the bounce of glycoforms ratio profiles. For example, monitoring the extension of mAb macro-heterogeneity or the ratio of glycosylated mAb with total mAb concentration, would add the error of predictions of both glycosylated mAb model and total mAb model. This fact could compromise the resolution of the final mAb glycosylation profiles and so further control strategies. Therefore, the capability of models to clearly show these final profiles was investigated in the form of a control chart as shown in Fig. 5.

Only best models were used for calculating the mAb macro- and micro-heterogeneity profiles. For the in-line mAb macro-heterogeneity profile, the fraction of glycosylated mAb was calculated as the ratio of glycosylated mAb (estimated by SVR) with total mAb concentration (estimated by LWR). Only SVR models were used for mAb micro-heterogeneity monitoring. For the in-line mAb micro-heterogeneity profiles, the fraction of either fucosylated, galactosylated, sialylated and high mannose glycoforms was calculated as the ratio of particular glycoform with glycosylated mAb concentration. Then micro-heterogeneity profiles could be monitored in real time as the fraction of glycoforms containing particular sugar moieties within the glycosylated

chain, which are closely related to clinical properties of mAb medicine.

Once mAb concentration was higher than  $30 \text{ mg.L}^{-1}$ , models allowed proper monitoring of mAb glycosylation profiles. This approach was encouraging for monitoring mAb macro-heterogeneity since accurate tendencies were observed during the whole culture, particularly for the abrupt decrease of mAb glycosylation around 100 h after the beginning of the process. As for mAb micro-heterogeneity, particularly for the fucosylated glycoform profile, a limited capacity was observed between 45 h–70 h of the culture. This behaviour was also observed, though to a lesser extent, for the galactosylated glycoform fraction profile. On the other hand, sialylated and high mannose isoforms profiles were properly estimated even at low concentrations. Results demonstrated the potential of SVR, LWR and NIR spectroscopy for real time monitoring of mAb glycosylation properties during CHO cell culture processes. Moreover, accuracy on concentration monitoring also permitted monitoring of accurate trends of mAb glycoforms ratios, closely related to mAb clinical effects. Then such mAb glycoforms ratios could be used as target values for later control.

Overall, these are encouraging results for the use of NIR spectroscopy for developing new retro-control systems. However, caution must be taken when discussing eventual prediction capability of LWR and SVR in particular for mAb production processes. The same consideration should be shown in the case of NIR spectroscopy, as it contains both physical and chemical information linked to chemical and physical phenomena within processes. There is always the possibility of new variables in new production processes that have not been considered yet are critical to the performance of this newly developed prediction platform.

## 5. Conclusions

Data demonstrating the feasibility of NIRS to monitor mAb glycosylation *in situ* have been presented. In this study, the monitoring of both macro- and micro-heterogeneity of glycosylated mAb was improved by the novel use of sample space-based regression methods, particularly SVR, that could handle non-linear relationships between glycoforms and spectra. As far as it can be asserted, this is the first report of real-time and *in situ* monitoring of mAb macro- and micro-heterogeneity using NIR spectroscopy as well as the first report of LWR and SVR methods for cell culture monitoring. Such methods dealing not only with chemical but also some physical information contained within spectra, highlight the importance of considering the strongly dynamic nature of cell culture processes for accurate monitoring by calibration models.

There is an increasing number of new mAb producing processes including mAb biosimilars and biobetters [35], and one can assume that *in situ* spectroscopy methods will be implemented systematically to fulfil the demand in regard to quality. This study lays the foundation for future studies to expand the capabilities of *in situ* spectroscopy and multivariate analysis to monitor mAb properties so that enhanced retro-control strategies can be established, leading to a more efficient design and control of processes using PAT and the Quality by Design principles.

## Declaration of Competing Interest

The authors declare no financial or commercial conflict of interest.

## Acknowledgements

The authors acknowledge the agencies that granted 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 an Eiffel Excellence scholarship. They also acknowledge the French National Agency of the Research (ANR) for supporting the ProCell-In-Line project.

## References

- [1] A.P. Teixeira, R. Oliveira, P.M. Alves, M.J.T. Carrondo, Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative, *Biotechnol. Adv.* 27 (2009) 726–732, <https://doi.org/10.1016/j.biotechadv.2009.05.003>.
- [2] S. Sinha, G. Pipes, E.M. Topp, P.V. Bondarenko, M.J. Treuheit, H.S. Gadgil, Comparison of LC and LC/MS methods for quantifying N-glycosylation in recombinant IgGs, *J. Am. Soc. Mass Spectrom.* 19 (2008) 1643–1654, <https://doi.org/10.1016/j.jasms.2008.07.004>.
- [3] A.S. Rathore, H. Winkle, Quality by design for biopharmaceuticals, *Nat. Biotechnol.* 27 (2009) 26–34, <https://doi.org/10.1038/nbt0109-26>.
- [4] J. Dong, N. Migliore, S.J. Mehrman, J. Cunningham, M.J. Lewis, P. Hu, High-throughput, automated protein A purification platform with multiattribute LC–MS Analysis for advanced cell culture process monitoring, *Anal. Chem.* 88 (2016) 8673–8679, <https://doi.org/10.1021/acs.analchem.6b01956>.
- [5] I. Burnina, E. Hoyt, H. Lynaugh, H. Li, B. Gong, A cost-effective plate-based sample preparation for antibody N-glycan analysis, *J. Chromatogr. A* 1307 (2013) 201–206, <https://doi.org/10.1016/j.chroma.2013.07.104>.
- [6] M. Doherty, J. Bones, N. McLoughlin, J.E. Telford, B. Harmon, M.R. DeFelippis, P.M. Rudd, An automated robotic platform for rapid profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture, *Anal. Biochem.* 442 (2013) 10–18, <https://doi.org/10.1016/j.ab.2013.07.005>.
- [7] A. Henninot, A. Terrier, J. Charton, R. Urbain, A. Fontayne, B. Deprez, T. Beghyn, Characterization of monoclonal antibodies by a fast and easy liquid chromatography–mass spectrometry time-of-flight analysis on culture supernatant, *Anal. Biochem.* 491 (2015) 52–54, <https://doi.org/10.1016/j.ab.2015.08.006>.
- [8] R. Moschetti, R. Massantini, M. Fidaleo, Application on-line NIR spectroscopy and other process analytical technology tools to the characterization of soy sauce desalting by electroanalysis, *J. Food Eng.* 263 (2019) 243–252, <https://doi.org/10.1016/j.jfoodeng.2019.06.022>.
- [9] M. Li, B. Ebel, F. Courtès, E. Guedon, A. Marc, In situ infrared spectroscopy as a PAT tool of great promise for real-time monitoring of animal cell culture processes, *Austin J. Anal. Pharm. Chem.* 3 (2016) 1–5.
- [10] S.A. Arnold, J. Crowley, N. Woods, L.M. Harvey, B. McNeil, In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation, *Biotechnol. Bioeng.* 84 (2003) 13–19, <https://doi.org/10.1002/bit.10738>.
- [11] R.A. Mattes, D. Root, D. Chang, M. Molony, M. Ong, In situ monitoring of CHO cell culture medium using near-infrared spectroscopy, *Bioprocess Int.* (2007) 46–50.
- [12] J.G. Henriques, S. Buziol, E. Stocker, A. Voogd, J.C. Menezes, Monitoring mammalian cell cultivations for monoclonal antibody production using near-infrared spectroscopy, in: G. Rao (Ed.), *Opt. Sens. Syst. Biotechnol.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 29–72, [https://doi.org/10.1007/10\\_2009\\_11](https://doi.org/10.1007/10_2009_11).
- [13] M. Clavaud, Y. Roggo, R. Von Daeniken, A. Liebler, J.-O. Schwabe, Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: prediction of multiple cultivation variables, *Talanta* 111 (2013) 28–38, <https://doi.org/10.1016/j.talanta.2013.03.044>.
- [14] H. Bhatia, H. Mehdizadeh, D. Drapeau, S. Yoon, In-line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models, *Eng. Life Sci.* 18 (2018) 55–61, <https://doi.org/10.1002/elsc.201700084>.
- [15] M.-Y. Li, B. Ebel, C. Paris, F. Chauchard, E. Guedon, A. Marc, Real-time monitoring of antibody glycosylation site occupancy by *in situ* Raman spectroscopy during bioreactor CHO cell cultures, *Biotechnol. Prog.* 34 (2018) 486–493, <https://doi.org/10.1002/btpr.2604>.
- [16] M. Li, B. Ebel, F. Chauchard, E. Guédon, A. Marc, Parallel comparison of *in situ* Raman and NIR spectroscopies to simultaneously measure multiple variables toward real-time monitoring of CHO cell bioreactor cultures, *Biochem. Eng. J.* 137 (2018) 205–213, <https://doi.org/10.1016/j.bej.2018.06.005>.
- [17] M. Schiestl, T. Stangler, C. Torella, T. Čepeljnik, H. Toll, R. Grau, Acceptable changes in quality attributes of glycosylated biopharmaceuticals, *Nat. Biotechnol.* 29 (2011) 310–312, <https://doi.org/10.1038/nbt.1839>.
- [18] I.J. del Val, C. Kontoravdi, J.M. Nagy, Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns, *Biotechnol. Prog.* 26 (2010) 1505–1527, <https://doi.org/10.1002/btpr.470>.
- [19] R.A. Dwek, A.C. Lellouch, M.R. Wormald, Glycobiology: “the function of sugar in the IgG molecule”, *J. Anat.* 187 (1995) 279–292.
- [20] A. Höskuldsson, PLS regression methods, *J. Chemom.* 2 (1988) 211–228, <https://doi.org/10.1002/cem.1180020306>.
- [21] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *TrAC Trends Anal. Chem.* 28 (2009) 1201–1222, <https://doi.org/10.1016/j.trac.2009.07.007>.
- [22] J. Huang, S. Romero-Torres, M. Moshgbar, Practical considerations in data pre-treatment for NIR and Raman spectroscopy, *Am. Pharm. Rev.* 13 (2010) 116–127.
- [23] W.S. Cleveland, S.J. Devlin, Locally weighted regression: an approach to regression analysis by local fitting, *J. Am. Stat. Assoc.* 83 (1988) 596–610, <https://doi.org/10.1080/01621459.1988.10478639>.
- [24] T. Næs, T. Isaksson, Locally weighted regression in diffuse near-infrared transmittance spectroscopy, *Appl. Spectrosc.* 46 (1992) 34–43, <https://doi.org/10.1366/0003702924444344>.
- [25] T. Næs (Ed.), *A User-Friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, UK, 2004.
- [26] Tormod Naes, Tomas Isaksson, Bruce Kowalski, Locally weighted regression and scatter correction for near-infrared reflectance data, *Anal. Chem.* 62 (1990) 664–673, <https://doi.org/10.1021/ac00206a003>.
- [27] European Medicine Agency, Guideline on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry and the Data Requirements for New Submissions and Variations, (2014) (Accessed November 10, 2019), [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-use-near-infrared-spectroscopy-pharmaceutical-industry-data-requirements-new-submissions\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-use-near-infrared-spectroscopy-pharmaceutical-industry-data-requirements-new-submissions_en.pdf).
- [28] U.S. Food & Drug Administration, Development and Submission of Near Infrared Analytical Procedures Guidance for Industry, (2015) (Accessed November 10, 2019), <https://www.fda.gov/media/91343/download>.
- [29] R.P. Cogdill, P. Dardenne, Least-squares support vector machines for chemometrics: an introduction and evaluation, *J. Infrared Spectrosc.* 12 (2004) 93–100, <https://doi.org/10.1255/jnirs.412>.
- [30] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.* 14 (2004) 199–222, <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.
- [31] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer New York, New York, NY, 2000, <https://doi.org/10.1007/978-1-4757-3264-1>.
- [32] N.J. Nilsson, *Learning Machines: Foundations of Trainable Pattern-Classifying Systems*, McGraw-Hill, 1965, <https://books.google.com.mx/books?id=7ENDAAAIAAJ>.
- [33] M. Awad, R. Khanna, *Support vector regression*, *Effic. Learn. Mach.* Apress, Berkeley, CA, 2015, pp. 67–80, [https://doi.org/10.1007/978-1-4302-5990-9\\_4](https://doi.org/10.1007/978-1-4302-5990-9_4).
- [34] H. Martens, J.P. Nielsen, S.B. Engelsen, Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures, *Anal. Chem.* 75 (2003) 394–404, <https://doi.org/10.1021/ac020194w>.
- [35] I.J. del Val, P.M. Jedrzejewski, K. Exley, S. Nga, S. Kyriakopoulos, K.M. Polizzi, C. Kontoravdi, Application of quality by design paradigm to the manufacture of protein therapeutics, in: S. Petrescu (Ed.), *Glycosylation*, InTech, 2012, <https://doi.org/10.5772/50261>.