



Article

Monitoring of Biopolymer Production Process Using Soft Sensors Based on Off-Gas Composition Analysis and Capacitance Measurement †

Pavel Hrnčířík

Department of Computing and Control Engineering, University of Chemistry and Technology Prague, Technická 5, 16628 Praha, Czech Republic; pavel.hrnccirik@vscht.cz

† This article is an expanded and updated version of the conference paper: Hrnčířík, P. Software sensors for monitoring of biopolymer production. In Proceedings of the 23rd International Conference on Process Control (PC), Virtual Event, Štrbské Pleso, Slovakia, 1–4 June 2021, pp. 308–312.

Abstract: This paper focuses on the design of soft sensors for on-line monitoring of the biotechnological process of biopolymer production, in which biopolymers are accumulated in bacteria as an intracellular energy storage material. The proposed soft sensors for on-line estimation of the biopolymer concentration represent an interesting alternative to the traditional off-line analytical techniques of limited applicability for real-time process control. Due to the complexity of biochemical reactions, which make it difficult to create reasonably complex first-principle mathematical models, a data-driven approach to the design of soft sensors has been chosen in the presented study. Thus, regression methods were used in this design, including multivariate statistical methods (PLS, PCR). This approach enabled the creation of soft sensors using historical process data from fed-batch cultivations of the *Pseudomonas putida* KT2442 strain used for the production of medium-chain-length polyhydroxyalkanoates (mcl-PHAs). Specifically, data from on-line measurements of off-gas composition analysis and culture medium capacitance were used as input to the soft sensors. The resulting soft sensors allow not only on-line estimation of the biopolymer concentration, but also the concentration of the cell biomass of the production bacterial culture. For most of these soft sensors, the estimation error did not exceed 5% of the measurement range. In addition, soft sensors based on capacitance measurement were able to accurately detect the end of the production phase. This study thus offers an innovative and practically relevant contribution to the field of monitoring of bioprocesses used for the production of medium-chain-length biopolymers.

Keywords: soft sensors; process monitoring; biopolymers; PLS and PCR regression; capacitance measurement; off-gas composition analysis



Citation: Hrnčířík, P. Monitoring of Biopolymer Production Process Using Soft Sensors Based on Off-Gas Composition Analysis and Capacitance Measurement. *Fermentation* **2021**, *7*, 318. <https://doi.org/10.3390/fermentation7040318>

Academic Editor: Cristiano Varrone

Received: 28 November 2021

Accepted: 14 December 2021

Published: 18 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The range of applications of soft sensors in the monitoring of industrial production biotechnological processes is increasing proportionally with the increasing demands on the quality of the production processes and the final products. In comparison to costly and relatively complex analytical techniques, soft sensors can be a more advantageous solution, in particular for the monitoring of those bioprocesses which are operated as fed-batch cultivations. These cultivations are typical for their complex reaction kinetics. Additional variability is due to variable composition of substrates and frequent changes in the production bioprocesses, as a result of the production of many different products. Therefore, soft sensors can be applied not only for the monitoring of the process itself, but additionally also for the evaluation of the quality of cultivation substrates and microbial cultures used as seed [1–3].

The underlying principle of soft sensors is to use a set of more or less commonly on-line measurable process variables for the estimation of other process variables or process

indicators that are difficult to measure on-line or can only be measured with very low sampling times. In general, it is possible to distinguish between two elementary types of soft sensors [4,5], as follows:

- “gray box” type of soft sensor, alternatively referred to as “model-driven”—using a first-principle mathematical model of the monitored process based on physical, chemical, or biological relationships with experimental identification of unknown parameters from historical process data;
- “black box” type of soft sensor, alternatively referred to as a “data-driven”—where the mathematical model describing the relationship between the inputs and outputs of the soft sensor is not known in advance, and hence the mathematical description of this relationship must be designed on the basis of historical process data using suitable computational tools, e.g., regression analysis, neural networks, etc.

Soft sensors of the “gray box” type that are used in the area of biotechnological processes are typically based on mass or energy balances, usually in combination with models of kinetic relationships describing the growth of the microbial culture. The calculation of the soft sensor outputs itself is then realized using one of the traditional estimation algorithms, such as Kalman filter or extended Kalman filter. The main problem in using this type of soft sensor for the monitoring of bioprocesses is a considerable difficulty associated with the process of designing mathematical models that could describe the monitored process with sufficient accuracy. For this reason, this type of sensor is not very widespread in the field of bioprocess monitoring [6]. In addition to the above-described complications associated with bioprocess modeling, the following typical properties of biotechnological processes also complicate the soft sensor design [5]:

- high sensitivity of the production microbial culture to changes in cultivation conditions such as pH, temperature, etc.;
- during the cultivation itself, the microbial culture passes through various physiological states, which usually result in different types of culture behavior;
- key parameters of bioprocess models typically change during cultivation, whereas on-line measurement, or at least estimation, of these changes is rather complicated.

For the above-outlined reasons, soft sensors of the second type (“black box” or “data-driven”) are more appropriate for the monitoring of biotechnological processes. Neural networks, fuzzy systems [7], various multivariate statistical methods [8], or other variants of regression analysis methods are used successfully for their design [5].

A key factor in the design of a soft sensor is not only the selection of a suitable calculation method, but also the selection of suitable inputs to the sensor in the form of on-line measured process variables. Soft sensors that use as their input on-line measurement of the composition of the gases leaving the bioreactor (the so-called off-gas) are relatively widespread [9]. These soft sensors are in their basic form frequently applied for on-line calculation of so-called derived process variables, such as oxygen uptake rate (*OUR*), CO₂ production rate (*CPR*), respiratory quotient (*RQ*), or oxygen transfer coefficient from gaseous to liquid phase in a bioreactor (*k_{La}*). More advanced soft sensors can be used for the on-line estimation of key bioprocess indicators, such as biomass concentration, biomass growth rate, or concentrations and rates of production of the main products (see, e.g., [10–14]).

This paper is part of a larger study that focuses on the design of data-driven soft sensors for on-line monitoring of the biotechnological process of mcl-PHAs biopolymer production. Specifically, this paper deals with the topic of data-driven soft sensors for estimating the concentrations of microbial biomass and biopolymer in the bioreactor. Two basic types of data inputs to the soft sensor are considered. First, attention is paid to soft sensors using on-line measurement of the composition of the off-gas leaving the bioreactor [15]. In the following part, attention is focused on soft sensors, which use as their input on-line capacitance measurement of the fermentation broth with the suspended microbial culture in the bioreactor. The reasons why the second part of the study focused

on examining the use of on-line capacitance measurement not only for on-line estimation of bacterial biomass concentration, but especially for on-line estimation of mcl-PHAs biopolymer concentration, were as follows: Capacitance measurement has in the past been successfully used for the on-line monitoring of lipid storage in yeast cells [16,17], as well as for the evaluation of the optimum harvesting time in the production of short-chain-length polyhydroxyalkanoates (scl-PHAs) [18,19]. The aim of this part of the study was therefore to examine the possibilities of using the capacitance measurement for on-line monitoring of the bacterial mcl-PHAs production processes.

2. Materials and Methods

2.1. Process Description

The fed-batch cultivation process of the bacterium *Pseudomonas putida* KT2442 used for the production of the mcl-PHAs biopolymer was chosen as a model process, for which a set of historical experimental process data was available. All cultivations used as a data source for this study were carried out in a 7-litre laboratory bioreactor at the Bioprocess Control Laboratory of the University of Chemistry and Technology in Prague. The experimental conditions for all cultivations were kept constant as follows: temperature 30 °C, pH 7, stirrer speed 900 1/min, aeration rate 9.5 L/min. After the initial batch phase carbon source (octanoic acid) was continually supplied to the bioreactor with a feeding rate set by the operator. Feeding strategies varied by individual cultivations, typically there was a phase of an exponential feeding followed by underfeeding and starvation, respectively. A detailed description of the individual measurements, which were the source of historical process data, is summarized in the following section.

2.2. Process Data Measurement and Analytical Methods

The oxygen and carbon dioxide concentrations in the off-gases exiting the bioreactor were measured by Servomex Servopro 1100A (O_2) and 1400B (CO_2) process analyzers (Spectris plc, Egham, UK) and these measurements were stored with a 1 min sampling period. The resulting off-gas composition measurement data were then used for on-line calculation of the *OUR* and *CPR* rates, as well as their corresponding cumulative values—cumulative oxygen consumption (*COC*) and cumulative carbon dioxide production (*CCP*), using Equations (1)–(4):

$$OUR = \frac{\dot{V}_A \rho_A}{V_L M_A} \left(\Delta O_2 - O_2 \left(\frac{N_2}{N_2 + \Delta O_2 - \Delta CO_2} - 1 \right) \right) M_{O_2} k_c \quad (1)$$

$$CPR = \frac{\dot{V}_A \rho_A}{V_L M_A} \left(\Delta CO_2 - CO_2 \left(\frac{N_2}{N_2 + \Delta O_2 - \Delta CO_2} - 1 \right) \right) M_{CO_2} k_c \quad (2)$$

$$COC = \int_0^t OUR(\tau) d\tau \quad (3)$$

$$CCP = \int_0^t CPR(\tau) d\tau \quad (4)$$

where \dot{V}_A (m^3/s) is the volumetric air flow rate on the inlet to the bioreactor, V_L (m^3) is the broth volume in the bioreactor, ρ_A (kg/m^3) is the air density, M_A (kg/mol) is the molecular weight of air, ΔO_2 (% vol.) is the difference between oxygen concentrations in the inlet air and the off-gas, ΔCO_2 (% vol.) is the difference between carbon dioxide concentrations in the inlet air and the off-gas, O_2 (% vol.) is the oxygen concentration in the off-gas, CO_2 (% vol.) is the carbon dioxide concentration in the off-gas, N_2 (% vol.) is the nitrogen concentration in the air (assumed to be constant at 79.07%), M_{O_2} (kg/mol) is the molecular weight of oxygen, M_{CO_2} (kg/mol) is the molecular weight of carbon dioxide, k_c is the coefficient for the conversion of concentration values from volume percent into

dimensionless volume fraction ($k_{conv} = 1/100 = 0.01$), t (min) is the current cultivation time, and τ (min) is the variable of integration (takes on values from time 0 to the current t).

The capacitance measurement of the fermentation broth with the suspended microbial culture in the bioreactor was performed with an Aber Biomass Monitor 210 (Aber Instruments Ltd., Aberystwyth, UK) operating in scanning mode and equipped with a four-annular ring probe AberProbe. The measuring device measured the capacitance spectrum (25 capacitance measurements corresponding to increasing frequencies on logarithmic scale from 0.1 MHz to 19.49 MHz) with a sampling period of 15 s. The measured spectral data were smoothed using a combination of two digital filters before further analysis. First, outliers were removed using a moving median filter with a window length of 60 data points (15 min) and the resulting signal was smoothed by a Savitzky-Golay filter with a third order polynomial fitting and a window length of 200 data points (50 min). These specific time window length settings were chosen according to the recommendations published in the study by Maskow et al. [16]. Alternative settings were tested, but they did not lead to better results. The capacitance data processed in this way were subsequently resampled using a sampling period of 1 min, corresponding to the sampling rate of the data from the measurement of the composition of the off-gases. See Figure 1 for a typical character of the filtered capacitance spectrum data.

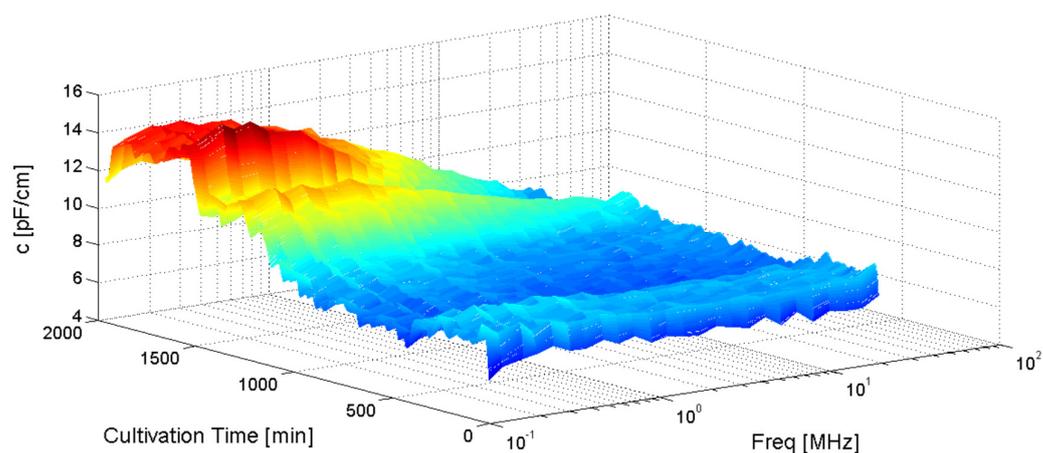


Figure 1. Sample of capacitance spectrum data.

Biomass concentration in the bioreactor was determined off-line gravimetrically as cell dry weight (CDW). Biopolymer concentration (PHA) was also determined off-line gravimetrically using an analytical procedure based on Soxhlet extraction. Samples for these two off-line analyses were taken at least every 3 h, in the final phase sampling was typically more frequent. The sampling times were chosen so that the samples captured key moments of the cultivation process, including transitions between the various process stages. To detect these transitions, the process operator used, among other indicators, on-line capacitance measurement provided by the Biomass Monitor 210 analyzer. In order to obtain a one-to-one correspondence between the on-line and the off-line measured data, the cell dry weight and biopolymer concentration data were interpolated using a piecewise cubic polynomial interpolation onto regular 1-min intervals. The combination of this interpolation method with off-line data measured at key moments of the cultivation process ensured that possible information losses due to interpolation were minimized. The resulting data set thus contained 4022 data records.

2.3. Soft Sensors Based on Off-Gas Analysis

Soft sensors for biomass concentration estimation based on the measured data describing the composition of the off-gas at the bioreactor outlet have been reported in several applications [8,20,21]. The rates of oxygen uptake (OUR) and carbon dioxide production (CPR) are closely linked to the biomass growth in microbial cultivation processes. In aerobic

microbial cultivation processes this relationship can be described by the Luedeking–Piret type of equation [20], (Equation (5)) for *OUR* and (Equation (6)) for *CPR*, respectively.

$$OUR = Y_{O/X} \frac{dc_X}{dt} + m_O c_X \quad (5)$$

$$CPR = Y_{C/X} \frac{dc_X}{dt} + m_C c_X \quad (6)$$

where $Y_{O/X}$ (kg oxygen consumed per 1 kg biomass produced) is the yield coefficient relating oxygen consumption to biomass production, c_X (kg/m³) is the microbial cell concentration in the bioreactor, t (min) is the cultivation time, m_O (kg oxygen consumed per 1 kg biomass per 1 min) is the oxygen consumption coefficient related to maintenance, $Y_{C/X}$ (kg carbon dioxide produced per 1 kg biomass produced) is the yield coefficient relating carbon dioxide production to biomass production, and m_C (kg carbon dioxide produced per 1 kg biomass per 1 min) is the carbon dioxide coefficient related to maintenance. If the yield coefficients can be assumed to be constant, and the contribution of the maintenance part of Equations (5) and (6) is low enough to be neglected, then the relationship between the cumulative (integral) values of O₂ consumption (*COC*) and CO₂ production (*CCP*), and the biomass concentration can be considered to be linear [20,22].

Compared to soft sensors for the estimation of biomass concentration, less attention has been paid to soft sensors for estimating product concentration (in the form of a storage substance-biopolymer) based on the on-line measurement of off-gas composition. The reason among others is a lower level of knowledge of related mechanisms. However, in cases where it is assumed that the biopolymer formation is closely related to the growth of microbial biomass, an analogous approach can be used as in the design of soft sensors for biomass concentration estimation. Consequently, in this study, the possibility of using both rates (*OUR*, *CPR*) and their cumulative values (*COC*, *CCP*) to estimate both the concentration of biomass and the concentration of biopolymer product using soft sensors was investigated.

2.4. Soft Sensors Based on Capacitance Measurement

Capacitance measurement provided by Aber Biomass Monitor type analyzers is standardly used in biotechnological processes to estimate the concentration of living microbial biomass in a bioreactor [16,17]. Not the entire measured spectrum is usually used for these purposes, but rather the difference between capacitance measurements (ΔC) at 2 selected frequencies according to Equation (7).

$$\Delta C = C(f_L) - C(f_H) \quad (7)$$

where $C(f_L)$ is capacitance measurement at low frequency, which reflects the concentration of living cells in the bioreactor, and $C(f_H)$ is background capacitance measurement at high frequency corresponding to high frequency capacitance contributions from non-cell components, such as air bubbles, etc. While f_H is typically chosen as the highest available frequency within the measured capacitance spectrum (e.g., 19.49 MHz in the case of Aber Biomass Monitor 210), the main frequency f_L is chosen so that the capacitance measurement at this particular frequency is best correlated with the live cell concentration (compared to capacitance measurement at other frequencies).

The results of several studies have shown that capacity measurements can be successfully used not only to estimate the concentration of living cells, but also to monitor the accumulation of lipid storage substances in yeast cells [16,17], as well as to detect the transition between production and consumption phases of some intracellular storage biopolymers (scl-PHAs) in bacterial cultures [18].

Following these findings, the presented study also focused on the possibility of using both two-capacity difference measurement (ΔC) and the capacitance spectra to estimate

the concentration of bacterial biomass and especially the concentration of the intracellular mcl-PHAs biopolymer.

3. Results and Discussion

3.1. Regression Analysis of the Relationship between the On-Line Process Data and Off-Line Concentrations Measurements

The first part of the regression analysis was focused on examining the relationship between on-line measured process variables calculated on the basis of the off-gas composition and both off-line variables, i.e., biomass and biopolymer concentrations. In total, four different variants of linear relationships between on-line variables and off-line determined concentrations of biomass and biopolymer were investigated. Specifically, the subject of the regression analysis was not only the relationship between cumulative variables (*COC*, *CCP*) and off-line concentrations, but also the relationship between the rates (*OUR*, *CPR*) and off-line concentrations. Both rates were also included in the regression analysis because studies documenting linearity of this relationship have been published for short chain biopolymers [23,24]. These four investigated variants were as follows:

- Case 1: biomass and biopolymer concentrations, respectively, are assumed to be linearly dependent on oxygen uptake rate *OUR*;
- Case 2: biomass and biopolymer concentrations, respectively, are assumed to be linearly dependent on carbon dioxide production rate *CPR*;
- Case 3: biomass and biopolymer concentrations, respectively, are assumed to be linearly dependent on cumulative oxygen consumption *COC*;
- Case 4: biomass and biopolymer concentrations, respectively, are assumed to be linearly dependent on cumulative carbon dioxide production *CCP*.

Historical experimental data from a set of fed-batch cultivations of the bacterium *Pseudomonas putida* KT2442 (see Sections 2.1 and 2.2) were used for the regression analysis. The results of the regression analysis in the form of values of coefficients of determination (R^2) for individual variants are summarized in Table 1. The overall statistical significance of the so obtained regression models has been tested and successfully verified using the standard F-test (at the 0.05 level).

Table 1. Results of regression analysis (on-line variables based on off-gas composition data vs. *CDW*, *PHA*).

| Variants | Coefficients of Determination (R^2) | |
|----------|---|---|
| | Biomass Concentration (<i>CDW</i>) | Biopolymer Concentration (<i>PHA</i>) |
| Case 1 | 0.46 | 0.40 |
| Case 2 | 0.59 | 0.53 |
| Case 3 | 0.98 | 0.98 |
| Case 4 | 0.97 | 0.98 |

It is apparent from the results of the regression analysis that in the case of both rates (*OUR*-case 1 and *CPR*-case 2) the assumption of linear dependence with both off-line concentrations is not justified. On the other hand, the assumption of a linear dependence of both cumulative quantities (*COC*-case 3 and *CCP*-case 4) and off-line concentrations was confirmed.

From the above, it can be concluded that in contrast to the biotechnological production process of short chain length biopolymer production, in the case of medium chain length biopolymer production process—which was the case in the presented study—the assumption of linear relationship between rates and off-line measured concentrations does not apply. On the contrary a strong linear relationship is observed between cumulative quantities and off-line concentrations. Furthermore, the theoretical assumption of significant mutual correlation of both cumulative variables—*COC* and *CCP*—was confirmed within the performed statistical analyses.

The second part of the regression analysis was then focused on examining the relationship between capacitance measurement differences (ΔC) calculated using Equation (7) and both off-line concentrations (CDW , PHA). Since the capacitance spectrum data are composed of 25 capacitance measurements corresponding to increasing frequencies on logarithmic scale from 0.1 MHz to 19.49 MHz, the calculation of the ΔC (using Equation (7)) corresponding to individual frequencies was performed in the following manner: The background capacitance measurement $C(f_H)$ was chosen, in line with the goal of removing high frequency capacitance contributions from non-cell components, such as air bubbles etc., as the capacitance measurement at the highest available frequency within the measured capacitance spectrum, i.e., $f_H = 19.49$ MHz. The main frequency f_L was variable and corresponded to the remaining 24 measuring frequencies (0.1, 0.12, ... 15.65 MHz). Hence, a total of 24 capacitance measurement differences ΔC_f (where $f = 0.1, 0.12, \dots 15.65$ MHz) were obtained. Each of the ΔC_f calculated in this way was then subjected to regression analysis as an independent variable in relation to both off-line concentrations CDW and PHA , respectively. The overall statistical significance of the so-obtained regression models has been tested and successfully verified using the standard F-test (at the 0.05 level). The resulting coefficients of determination (R^2) for individual variants are shown in the bar graphs in Figures 2 and 3.

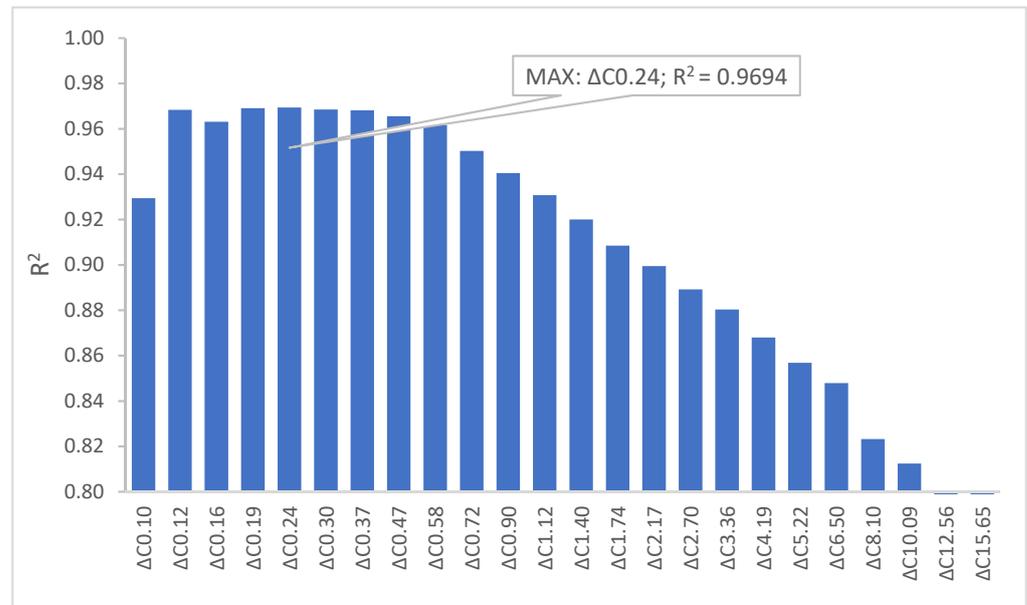


Figure 2. Coefficients of determination (R^2) of simple linear regressions-biopolymer concentration PHA as dependent variable, individual capacitance measurement differences ΔC_f as independent variables, respectively.

In the case of biopolymer concentration, the highest value of R^2 was reached at the capacitance measurement difference $\Delta C_{0.24}$, i.e., for the main frequency value $f_L = 0.24$ MHz. However, this maximum is not very distinct (see Figure 2); slightly lower values were achieved for other capacitance measurement differences, approximately in the range of main frequencies f_L from 0.12 to 0.58 MHz. In the case of biomass concentration, the highest value of R^2 was reached at the capacitance measurement difference $\Delta C_{0.58}$, i.e., for the main frequency value $f_L = 0.58$ MHz. Here, the maximum is more distinct than in the case of biopolymer concentration. However, even in this case the range of high values of R^2 corresponds approximately to the range of main frequencies f_L from 0.12 to 1.40 MHz, see Figure 3.

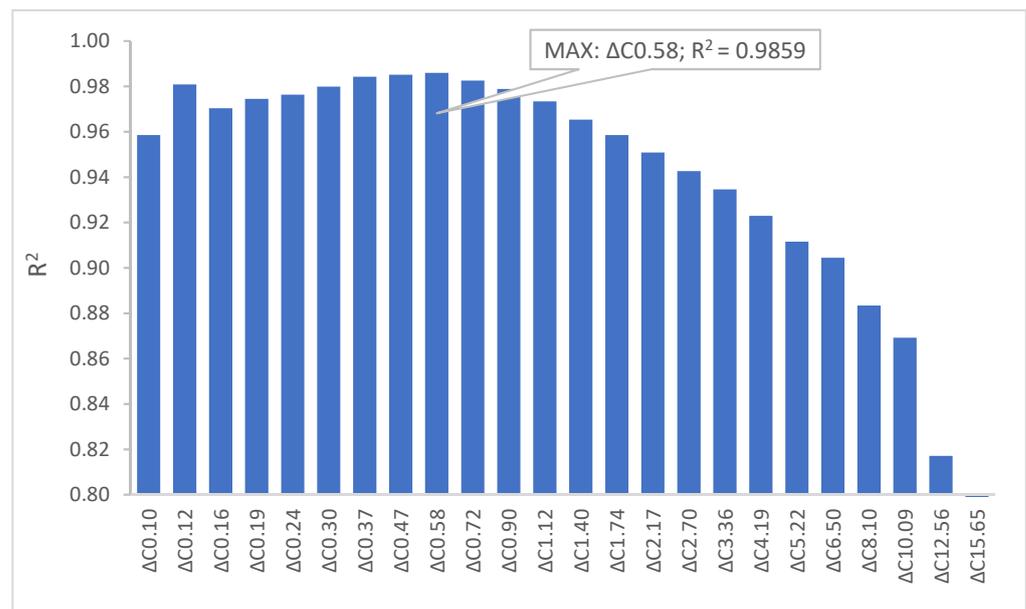


Figure 3. Coefficients of determination (R^2) of simple linear regressions—biomass concentration CDW as dependent variable, individual capacitance measurement differences ΔC_f as independent variables, respectively.

3.2. Design of Soft Sensors for Biomass and Biopolymer Concentration Estimation

Following the results of the regression analysis, two different types of soft sensors for biomass and biopolymer concentration estimation from the on-line measurement of off-gas composition and capacitance were proposed:

- Type 1: simply structured soft sensors based on simple linear regression, where biomass and biopolymer concentrations are estimated using only single input on-line process variable— COC or CCP or single ΔC_f , see Equation (8) and Figure 4a.

$$Out = k_1 Inp + k_0 \tag{8}$$

where Out is the output of the soft sensor, i.e., either the estimated biomass or biopolymer concentration, Inp is the input of the soft sensor, i.e., either COC or CCP or single selected ΔC_f (specifically $\Delta C_{0.24}$ for biopolymer and $\Delta C_{0.58}$ for biomass concentration in accordance with the above results of the regression analysis), and k_1 , k_0 are calibration constants of the specific soft sensor. This simple structure of the soft sensor (with only one independent variable at the sensor input) based on linear regression was chosen for two reasons—practical and statistical. The practical reason was that, in practice, both sets of on-line measurements (off-gas composition and capacitance measurements) are often not available at the same time, and even when only the off-gas composition is measured (in practice more widespread than the capacitance measurement), sometimes only one of the O_2 or CO_2 concentration measurements is available. The second, statistical reason why a soft sensor with multiple independent variables on the input based on multiple linear regression was not considered, was the fact that all considered variables (OUR , CPR , COC , CCP , ΔC_f) are mutually correlated and therefore the condition of real independence of input variables was not met. For this reason, multivariate regression methods based on partial least squares regression and principal component regression were used for a complex type of soft sensor with multiple inputs, see the following type 2.

- Type 2: comprehensively structured soft sensors based on multivariate statistical methods—partial least squares regression (PLS) or principal component regression (PCR), respectively, see Figure 4b. For a detailed description of both well-known methods, see, e.g., [25,26]. In both of these cases, the input of the soft sensors consisted

of both mutually correlated cumulative quantities based on off-gas composition measurement (*COC*, *CCP*) combined with a selected set of ΔC_f . The output of the sensors included both estimated concentrations (biomass and biopolymer). Specifically, the selected set of ΔC_f comprised 12 capacitance measurement differences ($\Delta C_{0.12}$, $\Delta C_{0.16}$, $\Delta C_{0.19}$, . . . , $\Delta C_{1.40}$), i.e., capacitance measurement differences corresponding to the range of main frequencies f_L from 0.12 to 1.40 MHz, for which high values of R^2 were attained in relation to off-line concentrations in the regression analyzes described in Section 3.1. Thus, there were a total of 14 on-line variables at the input of these type 2 soft sensors—*COC*, *CCP*, and 12 ΔC_f .

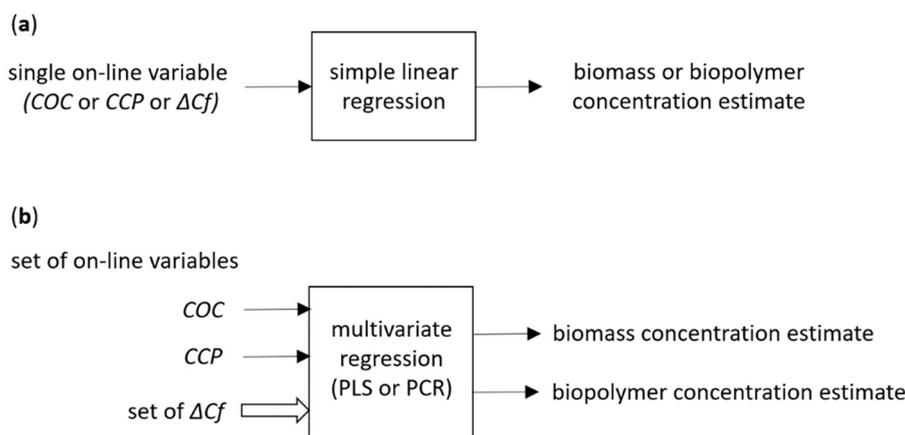


Figure 4. Input-output schemes of the proposed soft sensors-type 1 (a) and type 2 (b).

All proposed soft sensors were calibrated on the basis of the available historical experimental data using cross-validation. The data for the PLS and PCR-based sensors were additionally preprocessed using mean-centering. In the case of both multivariate methods, two latent variables (PLS) or principal components (PCR) were used, following findings from preliminary calculations in which this number of latent variables or principal components captured over 98% of variance.

Specifically, the variant of the cross-validation method used in this study was three-fold cross-validation. The dataset was hence randomly divided into three equal-sized subsamples. Of the three subsamples, a single subsample was retained as the validation data for testing, and the remaining two subsamples were used as training data. The cross-validation process was then repeated three times, with each of the three subsamples used just once as the validation data. For each round, a root-mean-square error of cross-validation (RMSE-CV) was calculated. The results were then averaged to produce a single estimation. The so obtained estimation of the sensor error are summarized in Table 2.

Table 2. Results of cross-validation.

| Sensor Types | RMSE-CV (% of Measurement Range) | |
|--------------------------------------|----------------------------------|-------------------------------------|
| | Biomass Concentration Estimation | Biopolymer Concentration Estimation |
| Type 1 (<i>COC</i> as input) | 3.64% | 4.63% |
| Type 1 (<i>CCP</i> as input) | 4.15% | 4.64% |
| Type 1 ($\Delta C_{0.24}$ as input) | – | 5.53% |
| Type 1 ($\Delta C_{0.58}$ as input) | 3.42% | – |
| Type 2 (PLS-based) | 2.78% | 4.46% |
| Type 2 (PCR-based) | 2.85% | 4.47% |

From the summary of the results in Table 2, it is evident that in the case of sensors for estimating the biopolymer concentration, the results obtained by the two proposed types of soft sensors are practically comparable. A simple type of sensor with one input (type 1) based on a simple linear regression thus provides similar estimation accuracy, in

comparison to a considerably more computationally complex sensor (type 2) based on multivariate statistical methods (PLS, PCR).

On the other hand, in the case of sensors for estimating biomass concentration, the results obtained by more complex types of sensors are slightly better (type 2 vs. type 1), but the differences are not very significant. From the comparison of the results obtained by the type 2 sensors based on multivariate statistical methods, the obtained results for PLS and PCR-based sensors are practically identical.

Special attention should be paid to the mutual comparison of simple soft sensors type 1 depending on the used input variable. From the perspective of data preprocessing, sensors using *COC* or *CCP* input are easier to design and use, as *COC* and *CCP* are cumulative quantities in the calculation of which integration is used (see Equations (3) and (4)). Hence, the measurement noise is significantly reduced to a satisfactory level, and no additional filtering is needed. On the other hand, soft sensors using capacitance measurement differences ΔC_f as their input are more complicated in that the capacitance measurement is usually marked by considerable noise (e.g., as a result of bioreactor aeration and other factors). The measured data must therefore be filtered by a combination of several digital filters before use in a soft sensor (see Section 2.2). However, sensors based on capacitance measurement differences are better able to capture the end of the biopolymer production phase at the end of cultivation, which is a key feature in terms of production monitoring (see Figures 5 and 6).

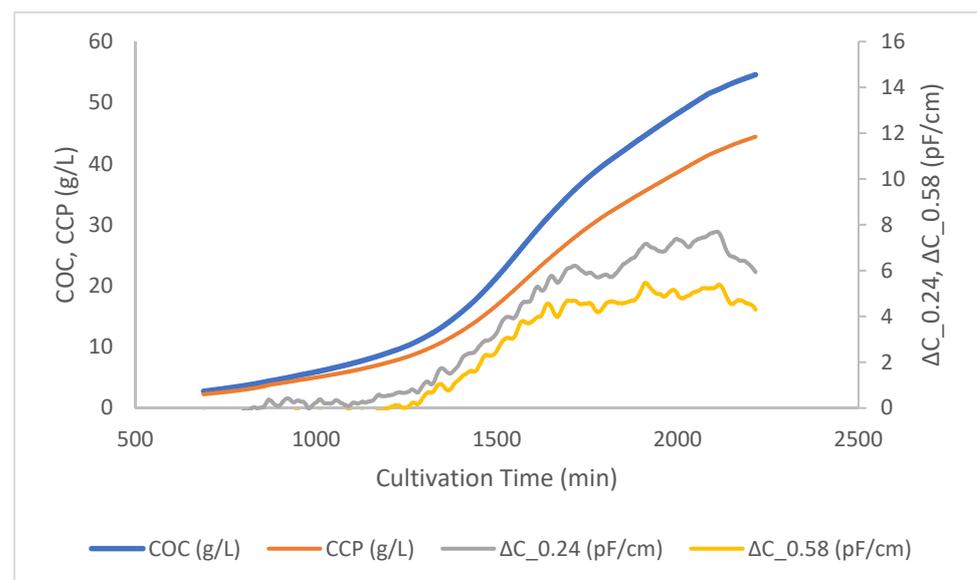


Figure 5. Model test cultivation—input on-line measured variables for soft sensors type 1.

Subsequently, five selected specific examples of soft sensors were tested—four soft sensors of the type 1 (two sensors for biomass with *COC* or $\Delta C_{0.58}$ as input; and two for biopolymer with *COC* or $\Delta C_{0.24}$ as input) and one complex soft sensor of the type 2 (PLS-based). On-line measured data from a typical biopolymer production cultivation with characteristic time course were used for testing (see Figures 5–7). From the output data obtained for the tested examples of soft sensors (Figures 6 and 7), it is clear that the differences between the tested sensors are minimal for most of the cultivation time. Larger differences are apparent towards the end phase of the production cultivation, when there is a significant slowing down of the cell culture growth and biopolymer production. In the case of both monitored concentrations (biomass and biopolymer), the final stagnation of growth, or even a decrease, is most evident from the outputs of the simple type 1 soft sensors, which have capacitance measurement differences at their input. It is therefore apparent also from this model example that soft sensors based on capacitance measurement can be very useful for early detection of the end of the production phase.

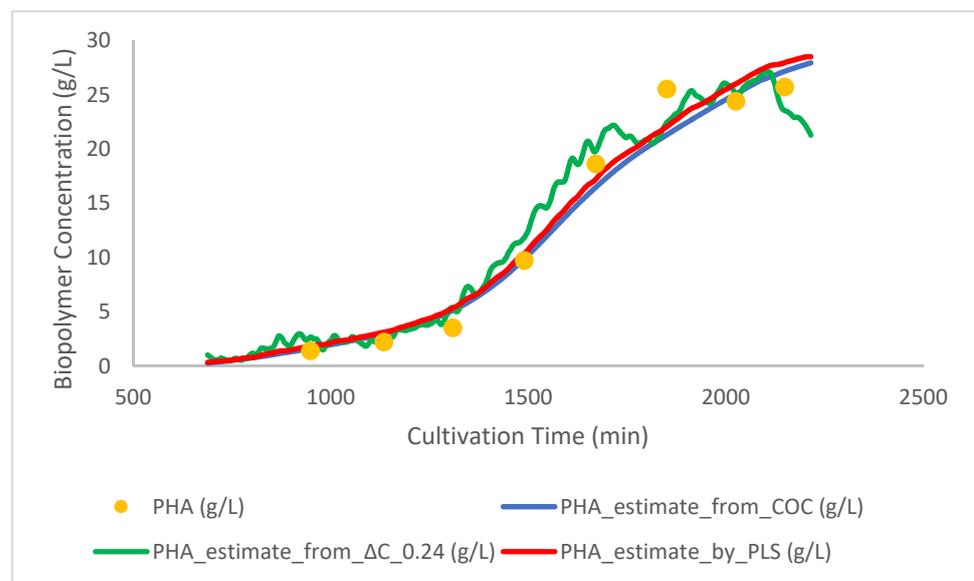


Figure 6. Model test cultivation-biopolymer concentration-comparison of off-line measurements and estimates using selected variants of soft sensors.

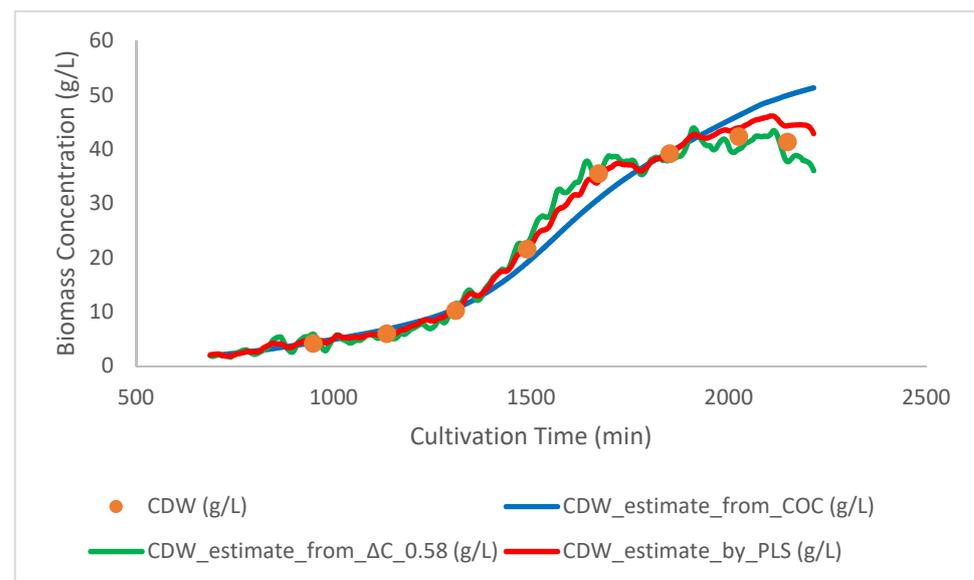


Figure 7. Model test cultivation-biomass concentration-comparison of off-line measurements and estimates using selected variants of soft sensors.

4. Conclusions

This paper presented a selection of soft sensors for estimating the concentrations of biomass and biopolymer in the mcl-PHAs biopolymer production process—from simple sensors based on linear regression to sensors based on statistical multivariate methods, such as PLS and PCR—which used on-line measurement of off-gas composition and capacitance measurement of the fermentation broth as input data. All types of the designed soft sensors were able to relatively accurately estimate both monitored concentrations during the growth phase of the process. Significant differences between the performance of individual soft sensors were apparent only in the final phase of the process. This phase is characterized by a significant slowing down of the microbial growth and biopolymer production. Only soft sensors based on capacitance measurement detected this important change in the trend of the cultivation process. On the other hand, this measurement is significantly affected by noise and must therefore be filtered in a complex way. For the

growth phase of the process, monitoring by soft sensors based on measurement of the composition of the off-gases (cumulative variables *COC* or *CCP*) is hence more suitable.

Further research will therefore be focused on the development of adaptive soft sensors that will be able to evaluate the quality of their estimation during the estimation process and continuously adjust the composition of their data inputs. The specific goal will be to design soft sensors that will use a different on-line measured variable or set of variables for each individual phase of the process.

Finally, a few more remarks on the limitations of the proposed approach to the design of soft sensors. The designed soft sensors were primarily intended for the monitoring of a specific bioprocess for the production of medium-chain-length biopolymers. However, the possibilities of using the proposed approach are wider. In general, this approach can find application in soft sensor design for the monitoring of all aerobic production bioprocesses where product formation is closely linked to biomass growth.

Funding: This research was funded by the institutional support RVO. This support is gratefully acknowledged.

Data Availability Statement: The data used in this study are property of the Bioprocess Control Laboratory of UCT Prague and are available on request.

Acknowledgments: The author thanks the staff of the Bioprocess Control Laboratory of UCT Prague for kindly providing historical process data from experimental cultivations.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Faergestad, E.M.; Oyaas, J.; Kohler, A.; Berg, T.; Naes, T. The use of spectroscopic measurements from full scale industrial production to achieve stable end product quality. *J. Food Sci. Technol.* **2011**, *44*, 2266–2272. [[CrossRef](#)]
2. Gao, Y.; Yuan, Y.J. Comprehensive quality evaluation of corn steep liquor in 2-keto-L-gulonate fermentation. *J. Agric. Food Chem.* **2011**, *59*, 9845–9853. [[CrossRef](#)]
3. Cunha, C.C.F.; Glassey, J.; Montague, G.A.; Albert, S.; Mohan, P. An assessment of seed quality and its influence on productivity estimation in an industrial antibiotic fermentation. *Biotechnol. Bioeng.* **2002**, *78*, 658–669. [[CrossRef](#)]
4. Luttmann, R.; Bracewell, D.G.; Cornelissen, G.; Gernaey, K.V.; Glassey, J.; Hass, V.; Kaiser, C.; Preusse, C.; Striedner, G.; Mandenius, C.F. Soft sensors in bioprocessing: A status report and recommendations. *Biotechnol. J.* **2012**, *7*, 1040–1048. [[CrossRef](#)]
5. Sharma, S.; Tamba, S.S. Softsensor development for biochemical systems using genetic programming. *Biochem. Eng. J.* **2014**, *85*, 89–100. [[CrossRef](#)]
6. Kadlec, P.; Gabrys, B.; Strandt, S. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.* **2009**, *33*, 795–814. [[CrossRef](#)]
7. Glassey, J.; Montague, G.A.; Ward, A.C.; Kara, B. Enhanced supervision of recombinant *E. coli* fermentations via artificial neural networks. *Proc. Biochem.* **1994**, *29*, 387–398. [[CrossRef](#)]
8. Ödman, P.; Johansen, C.L.; Olsson, L.; Gernaey, K.V.; Eliasson Lantz, A. On-line estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using in-situ multi-wavelength fluorescence and software sensors. *J. Biotechnol.* **2009**, *144*, 102–112. [[CrossRef](#)] [[PubMed](#)]
9. Aehle, M.; Kuprijanov, A.; Schaepe, S.; Simutis, R.; Luebbert, A. Simplified off-gas analyses in animal cell cultures for process monitoring and control purposes. *Biotechnol. Lett.* **2011**, *33*, 2103–2110. [[CrossRef](#)]
10. Chérury, A. Software sensors in bioprocess engineering. *J. Biotechnol.* **1997**, *52*, 193–199. [[CrossRef](#)]
11. Montague, G.A.; Morris, A.J.; Tham, M.T. Enhancing bioprocess operability with generic software sensors. *J. Biotechnol.* **1992**, *25*, 183–201. [[CrossRef](#)]
12. Ignova, M.; Glassey, J.; Ward, A.C.; Montague, G.A. Multivariate statistical methods in bioprocess fault detection and performance forecasting. *Trans. Inst. Meas. Control* **1997**, *19*, 271–279. [[CrossRef](#)]
13. Albiol, J.; Robustr, J.; Casas, C.; Poch, M. Biomass estimation in plant cell cultures using an extended Kalman filter. *Biotechnol. Prog.* **1993**, *9*, 174–178. [[CrossRef](#)]
14. Arnold, S.A.; Crowley, J.; Woods, N.; Harvey, M.L. In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation. *Biotechnol. Bioeng.* **2003**, *84*, 13–19. [[CrossRef](#)]
15. Hrnčířík, P. Software sensors for Monitoring of Biopolymer Production. In Proceedings of the 23rd International Conference on Process Control (PC), Virtual Event, Štrbské Pleso, Slovakia, 1–4 June 2021; pp. 308–312.
16. Maskow, T.; Röllich, A.; Fetzer, I.; Ackermann, J.U.; Harms, H. On-line monitoring of lipid storage in yeasts using impedance spectroscopy. *J. Biotechnol.* **2008**, *135*, 64–70. [[CrossRef](#)] [[PubMed](#)]

17. Maskow, T.; Röllich, A.; Fetzer, I.; Yao, J.; Harms, H. Observation of non-linear biomass-capacitance correlations: Reasons and implications for bioprocess control. *Biosens. Bioelectron.* **2008**, *24*, 123–128. [[CrossRef](#)]
18. Kedia, G.; Passanha, P.; Dinsdale, R.M.; Guwya, A.J.; Lee, M.; Esteves, S.R. Addressing the challenge of optimum polyhydroxyalkanoate harvesting: Monitoring real time process kinetics and biopolymer accumulation using dielectric spectroscopy. *Bioresour. Technol.* **2013**, *134*, 143–150. [[CrossRef](#)] [[PubMed](#)]
19. Li, L.; Wang, Z.J.; Chen, X.J.; Chu, J.; Zhuang, Y.P.; Zhang, S.L. Optimization of polyhydroxyalkanoates fermentations with on-line capacitance measurement. *Bioresour. Technol.* **2014**, *156*, 216–221. [[CrossRef](#)]
20. Jenzsch, M.; Simutis, R.; Eisbrenner, G.; Stückrath, I.; Lübbert, A. Estimation of biomass concentrations in fermentation processes for recombinant protein production. *Bioprocess Biosyst. Eng.* **2006**, *29*, 19–27. [[CrossRef](#)]
21. Sundström, H.; Enfors, S. Software sensors for fermentation processes. *Bioprocess Biosyst. Eng.* **2005**, *31*, 145–152. [[CrossRef](#)]
22. Rønnest, N.P.; Stocks, S.M.; Eliasson Lantz, A.; Gernaey, K.V. Introducing process analytical technology (PAT) in filamentous cultivation process development: Comparison of advanced online sensors for biomass measurement. *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 1679–1690. [[CrossRef](#)] [[PubMed](#)]
23. Liu, D.; Zhang, X.T.; Zhang, D.J.; Zeng, S.W.; Lu, P.L. Polyhydroxyalkanoate (PHA) Synthesis by Glycerol-Based Mixed Culture and Its Relation with Oxygen Uptake Rate (OUR). *Huan Jing Ke Xue* **2016**, *37*, 3518–3523. [[PubMed](#)]
24. Zeng, S.W.; Wang, Z.Y.; Gao, J.; Liu, D.; Zhang, D.J.; Lu, P.L. On-Line Estimation for the Amount of Stored PHA in Activated Sludge Based on OUR-HPR Measurements. *Huan Jing Ke Xue* **2015**, *36*, 1713–1719. [[PubMed](#)]
25. Jolliffe, I.T. A Note on the Use of Principal Components in Regression. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1982**, *31*, 300–303. [[CrossRef](#)]
26. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]