Comparison of noninvasive, in-situ and external monitoring of microbial growth in fed-batch cultivations in Corynebacterium glutamicum

Lira-Parada, Pedro A.; Tuveri, Andrea; Seibold, Gerd M.; Bar, Nadav

Published in:
Biochemical Engineering Journal

Link to article, DOI:
10.1016/j.bej.2021.107989

Publication date:
2021

Document Version
Peer reviewed version

Citation (APA):
Graphical Abstract

Comparison of noninvasive, *in-situ* and external monitoring of microbial growth in fed-batch cultivations in *Corynebacterium glutamicum*

Pedro A. Lira-Parada, Andrea Tuveri, Gerd M. Seibold, Nadav Bar
Highlights

Comparison of noninvasive, *in-situ* and external monitoring of microbial growth in fed-batch cultivations in *Corynebacterium glutamicum*

Pedro A. Lira-Parada, Andrea Tuveri, Gerd M. Seibold, Nadav Bar

- Growth was monitored using five different technologies.
- The sensors were robust to common bio-process variables like agitation and aeration.
- Noise inherited in the sensors was successfully reduced using Gaussian filter.
- Gaussian filter facilitated the estimation of biomass growth rates from the signals.
Comparison of noninvasive, *in-situ* and external monitoring of microbial growth in fed-batch cultivations in *Corynebacterium glutamicum*

Pedro A. Lira-Parada\textsuperscript{a}, Andrea Tuveri\textsuperscript{a}, Gerd M. Seibold\textsuperscript{b}, Nadav Bar\textsuperscript{a,∗}

\textsuperscript{a}Department of Chemical Engineering, Norwegian University of Science and Technology, N-7491 Trondheim, Norway (e-mail: pedro.a.l.parada@ntnu.no, nadi.bar@ntnu.no)

\textsuperscript{b}Section for Synthetic Biology, Department of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark (e-mail: gesei@dtu.dk)

Abstract

Automated bioprocess systems rely on accurate online measurements of cellular growth. In the present study, six sensors representing five different technologies, namely near-infrared absorbance, external non-invasive light scattering, back scattering reflectance, dielectric spectroscopy and exhaust gas analysis were evaluated in online monitoring of growth of the microbial cell factory *Corynebacterium glutamicum*. The signal outputs were monitored during the stationary growth phase of the cells under different conditions of agitation and aeration, conditions that pose potential interference with the *in-situ* signals. The random noise of the measurements was filtered using a Gaussian filter and the time derivatives were computed to obtain information of the microbial growth rate. Surprisingly, the signals of most probes under fed-batch cultivations of *C. glutamicum* were highly correlated with the stirring, but were mostly unaffected by the airflow, despite the increase in air bubbles that can cause measurement interference. Taken together, we show that an online estimation of growth of *C. glutamicum* based on the different sensor signals was possible with linear and multiple linear regression, correcting for the effect of

∗Corresponding author
stirring, and an estimation of growth rate was made accurate with an implementation of a Gaussian filter.

**Keywords:** Microbial cell factories, biomass sensors, microbial culture, dielectric spectroscopy, back-scattering reflectance, external sensors, non-invasive monitoring.

---

1. **Conflict of Interest Statement**

The authors declare they have no conflict of interest.

2. **Introduction**

   Microbial cultures are multiple-input multiple-output non-linear systems. Understanding the dynamics of these processes requires monitoring of cellular growth. For this purpose, several robust and accurate analytical techniques based on offline methods are well established. However, automated modelling, control, and optimization of biochemical processes require high frequency, reliable measurements without the delay forced by offline techniques [1]. Monitoring and control of bioprocess require well defined and suitable online analytical techniques and data processing methods for the specific bioprocess application. The advantages and limitations of these techniques and methods must be well understood in order to monitor the appropriate process variables effectively [2]. In this study, we used *Corynebacterium glutamicum*, a Gram positive, rod shaped bacterium that is a versatile industrial workhorse, producing amino acids [3], organic acids [4], and further bulk and fine chemicals [5, 6]. Currently, the bioprocess engineering community monitors growth in microbial cultures through different measurement principles and available technologies, both offline and online methods. The offline optical density measurement with spectrophotometer (OD\textsubscript{600}) and the determination of cell dry weight (CDW) are the most common
methods to estimate biomass concentration in a *C. glutamicum* culture [7]. However, these methods present a significant time delay (minutes to hours) between the acquisition of the samples and the estimation of the biomass values, and thus are not useful in automatic control of microbial system. Moreover, quantification by CDW measurements is particularly slow, as it requires filtration, weighing and drying process posing a major drawback for control applications [8].

It has been shown that multiple measurements can provide valuable information of the culture conditions [9]. In this study, five different sensor technologies were explored, namely: 1) Near Infrared Absorbance (NIR), 2) External light scattering source, 3) Capacitance, 4) Back scattering reflectance, and 5) An offline signal of CO\(_2\) off-gas measurement. The NIR spectroscopy has been already applied in online and in-line monitoring [10]. Previous studies involved quantification of substrate and product formation with R-squared \((R^2)\) values between 0.95 and 0.93 for biomass calibration [11], and growth monitoring in a 10 mL-scale bioreactor, with \(R^2\) between 0.98 and 0.91 in *C. glutamicum* [12]. Importantly, previous studies have also shown that biomass measurements are sensitive to impurities in the substrate mixture, such as the one exist in industrial residual sugars, and the choice of the biomass probe is highly important in such conditions [13]. A wide spectrum NIR in fed-batch cultivations of *Saccharomyces cerevisiae* demonstrated that the spectra are corrupted by disturbances (e.g. gas bubbles from sparging), which can result in noisy and defective spectra, potentially leading to highly imprecise OD measurements [14]. An external and non-invasive light scattering sensor has been implemented in yeast cultures, where the regression between the signal and CDW was \(R^2 = 0.98\) [15]. In microbial systems such as *Bacillus subtilis* [16, 17] and *Escherichia coli* [18], signals of the light scattering sensors were used to compute differences in
microbial growth rates, but these studies did not report the correlation with OD or CDW and their accuracy was not evaluated. The back-scattering reflectance has been studied in filamentous microorganisms to monitor the bioprocess, but reports on accuracy with respect to CDW or OD was not reported [19]. In online control applications of microalgae, the relative signal of this sensor was used to estimate the growth of biomass, but estimation of the growth rate exhibited high noise levels [20]. The capacitance technology has been evaluated in microbial growth rate control of filamentous fungi processes, where linear and multivariate models estimated the biomass concentration and the growth rate with $R^2$ values of viable cells between 0.87 and 0.57 [21] [22]. Additionally, in E. coli cultures, this sensor signal was correlated with biomass concentrations [23]. However, the signal deviated during the bio-process from linear correlation and presented $R^2$ values between 0.96 and 0.86. The CO$_2$ information is relevant in monitoring respiration and understanding the cellular metabolic processes [24]. Moreover, in yeast cultures, dissolved carbon dioxide strongly affected cellular growth [25]. Several kinetic studies have been dedicated to CO$_2$ modelling and monitoring, [26] [13] in order to estimate the growth of the culture. However, these measurements can only estimate the growth indirectly with a varying noise level, depending on the bioprocess phase (e.g. batch, feeding) [13]. Thus, there is a need for reliable in-situ measurements that can measure growth in a precise and rapid manner.

In this study, six different sensors, representing the five technologies mentioned above, were evaluated and compared with offline optical density measurements in C. glutamicum cultures. We monitored the sensors in steady state and applied different stirring rates and air flow conditions. The main objective was to understand the advantages and limitations of the different sensors in obtaining in-situ growth and growth rate estimates. The present study focuses
only on growth, and thus product formation was not evaluated.

3. Materials and Methods

*Corynebacterium glutamicum* WT strain ATCC13032 was used in this study. For growth experiments the bacteria were freshly streaked out from glycerol stocks on LB agar plates and incubated at 30 °C. Thereby obtained single colonies were used to inoculate 50 mL precultures in 2TY medium in 500 mL baffled shake flasks and incubated at 30°C and 200 rpm. The bacteria were separated from the 2TY broth by centrifugation at (4000 g, 10 min), the cell pellet was washed twice with saline (0.9 g/L NaCl), then re-suspended in 75 mL and used for the inoculation in the bioreactor.

3.1. Bioreactor procedure

The bioreactor contained 75 % v/v CGXII medium without MOPS, 20 % v/v glucose carbon source solution (100 g L⁻¹ glucose) and a 5 % inoculum v/v. The initial optical density was set to one and the initial glucose concentration was 20 g L⁻¹ from a source of 100 g L⁻¹. The experimental set-up (Figure 1) has Lucullus (Securecell, Switzerland) and EVE bioprocess platform device (Infors, Switzerland) as process information management system (PIMS) that monitored and recorded the signals. The pH was maintained at 7.0 by online measurement using an Hamilton EasyFerm Plus Arc 325 mm pH electrode (Hamilton, Switzerland) and addition of base (4 M KOH) and acid (phosphoric acid (10 % w/v)). The Anti-foam (AF) solution Antifoam 204 (Sigma Aldrich) was added manually to the bioreactor when required. The heating/cooling element consisted of a glass double jacket that preserved the temperature (T) of the system at 30°C. The relative partial oxygen pressure (pO₂) was measured online using a polarimetric oxygen electrode Hamilton VisiFerm DO Arc 325 mm (Hamilton, Switzerland), and the pO₂ values were adjusted above 30 % in
a cascade by stirring at 200 to 1100 rpm using a Rushton type impeller. A mass flow controller (MFC) supplied and controlled a constant 2 L min$^{-1}$ air flow. The optical density was measured using a UV-VIS spectrophotometer (UV-160A; Shimadzu). The results presented in this study are from six independent growth experiments with the same experimental conditions. Technically, it was not possible to set the different sensors in one single reactor, because bioreactor vendors restrict the number of fittings in the metallic top plate of the bioreactor.

3.2. Sampling method

The Numera system (Securecell, Switzerland) conducted the sampling from the bioreactors. The Numera system was programmed using Lucullus (Securecell, Switzerland). Previous studies with the Numera system showed that the dilution procedure has no impact on cell viability and cell debris, the monitoring of various analytes does not have significant errors, and the accuracy of the Numera system was one order of magnitude better than in line methods such as Raman spectroscopy [29]. The hardware consists of a multiplexer, a dilution module and a filtration unit (Figure 1). The multiplexer sampled culture broth from the different bioreactors and the samples were transferred into a dilution module, where the samples were diluted in a defined ratio of 1:10 with HPLC grade water. The samples were injected into vials by an auto-sampler at 5° C. In this study, an optical density was considered during the measurements, and it was observed in dedicated experiments that the optical density and the cell dry weight follow a linear relationship (See Appendix Figure A.10). To compare the OD measurements with the sensor signals, we used the NUMERA system to collect samples of 1 mL at desired times, and used this samples to measure OD or CDW manually, as described in [29]. Cell dry weight values were obtained after manually filtering 3 mL of culture broth in pre-weighted filters of 0.2 µm, the samples were rinsed with water and dried at 60° C for 72 h.
3.3. Sugars quantification

Sugars HPLC procedure was implemented using the Thermo Scientific™ Dionex™ Chromeleon™ Chromatography Data System, and the sugar chromatography was conducted using the Thermo Scientific Ulti Mate 3000 with a VA 300/7.8 NUCLEOGEL® Sugar Pb column. The mobile phase was HPLC water with a flow rate of 0.4 mL min⁻¹ HPLC, a pressure of 25 bar, and at a temperature 80 °C. The refractive index (RI) detector RefractoMax520 ERC module monitored the signals.

3.4. Sensors technology

The Near-infrared (NIR) absorption was measured using ASD12-N Absorption probes (Optek, Esser, Germany), with two different hardware configurations: OPL01 and OPL05. The sensors were set on the top plate of the bioreactor, and the data was acquired every 10 seconds. The external and non-invasive light scattering (LS) technology was monitored with the cell growth quantifier Bior(CGQ BIOR, Aquila biolabs GmbH, Germany). The LS external sensor was attached to the glass wall of the bioreactor, measuring with a controlled LED array using 520 nm wavelength and a photo-diode detector, recording the measurements every 30 minutes. The back-scattering (BS) reflectance was measured using the probe 3000 Fiber Optic Total Biomass sensor with a wavelength of 1310 nm every 60 seconds. The dielectric spectroscopy was applied using the Futura capacitance sensor (Aber Instruments, Aberstwyth, UK), and the data was acquired every 30 seconds. The exhaust gas concentration was measured every 10 seconds with the BlueInOne Ferm sensor connected in the gas outlet of the bioreactor. The infrared (IR) sensor was used for the CO₂, and a Zirconium Dioxide (ZrO₂) sensor was used for the O₂ gas phase quantification. The sensors provided on-line data that followed changes in optical density, the capacitance sensor is capable to detect viable cells [30], however, this was not explored
in this study. Perturbations of the stirring and air flow were programmed in Lucullus (Securecell, Switzerland), and the observations in Figures 4, 5 and 6 were obtained from the analysis of each signal output (see Figure A.12 in the appendix).

4. Theory/Calculation

4.1. General Bioreactor Model

Different nonlinear forms are available to model microbial processes [31]. In a general non-linear formulation, the model has a vector of $x(t)$ states with $x_0$ initial conditions, $u(t)$ inputs, $p$ parameters, time $t$, $y(t)$ observations and an output function $g$:

$$\dot{x}(t) = f(x(t), u(t), p)$$
$$y(t) = g(x(t), p)$$
$$x(t_0) = x_0$$

(1)

A fed-batch bioreactor mass balance considering the volume ($V$), biomass ($X$) and substrate ($S$) is [32]:

$$\dot{V} = F_{in}$$
$$\dot{X} = \frac{-XF_{in}}{V} + R_X$$
$$\dot{S} = -\frac{(S_{in} - S)F_{in}}{V} + R_S$$

(2) (3) (4)

The system has an inlet flow ($F_{in}$), a sugar concentration $S_{in}$, rate of biomass formation $R_X$ and rate of substrate consumption $R_S$. When a dilution term is
considered $D = \frac{E}{V}$, then Equations 3 and 4 become:

\[
\begin{align*}
\dot{X} &= -XD + R_X \quad (5) \\
\dot{S} &= -(S_{in} - S)D + R_S \quad (6)
\end{align*}
\]

The rate of biomass formation is a function of the microbial specific growth rate ($\mu$), and from Equation 5:

\[
\dot{X} = -XD + \mu X \quad (7)
\]

The rate of biomass formation has the following equality:

\[
\frac{1}{X} \frac{dX}{dt} = \frac{d\ln(X)}{dt} \quad (8)
\]

Then, Equation 7 becomes:

\[
\frac{d\ln(X)}{dt} = -D + \mu \quad (9)
\]

Assuming that the biomass concentration is proportional to the optical density ($X \propto OD$), then Equation 9 is:

\[
\frac{d\ln(X)}{dt} = \ln(\dot{OD}) = -D + \mu \quad (10)
\]

4.2. In-situ monitoring of Oxygen and Carbon Dioxide

The total CO$_2$ and O$_2$ were computed with the consideration that the O$_2$ inlet concentration was $y_{O_2-in} = 0.21$, and that there was no inlet flow of CO$_2$
(i.e. $y_{CO_2-in} = 0$) in the air-bottle:

\[
\dot{n}_{O_2}^{consumed} = \frac{PQ(y_{O_2} - 0.21)}{RT} \tag{11}
\]

\[
\dot{n}_{CO_2}^{produced} = \frac{PQy_{CO_2}}{RT} \tag{12}
\]

where $P$ is the pressure (1 atm), $Q$ is the gas flow (2 L·min$^{-1}$), $R$ is the ideal gas constant ($R = 0.082$ L·atm·mol$^{-1}$·K$^{-1}$), and $T$ is the temperature in Kelvin ($T = 303.15$ K). The total CO$_2$ and O$_2$ were estimated with the following integrals:

\[
n_{O_2,total} = \int_0^{t_{final}} \dot{n}_{O_2}^{consumed} dt \tag{13}
\]

\[
n_{CO_2,total} = \int_0^{t_{final}} \dot{n}_{CO_2}^{produced} dt \tag{14}
\]

### 4.3. In-situ Monitoring of Growth

In this study, optical density is derived from different in-situ signals. The measurement signals consist of different units and order of magnitudes. In order to enable an adequate comparison between the sensor signals, their values were subtracted with respect to their initial value and normalized using their maximum difference value. The normalization guarantees values between zero and one:

\[
Signal'(t_i) = \frac{Signal(t_i) - Signal_0}{\max_i[Signal(t_i) - Signal_0]} \tag{15}
\]

where the zero value is ($Signal_0$) and the prime indicates that the signal is normalized ($Signal'$). Similarly, the optical density obtained from the spec-
Trophotometer measurements was normalized by the following:

\[ OD'(t_i) = \frac{OD(t_i) - OD_0}{\max_i |OD(t_i) - OD_0|} \]  

(16)

where the zero value is \( OD_0 \), and the prime indicates normalized value \( OD'(t_i) \) with respect to their maximum difference values.

\[ E(p) = SSE = \sum_i (y_{exp-i}(t_i) - y'(t_i, p))^2 = \sum_i (OD'_{600}(t_i) - y'(t_i, p))^2 \]  

(18)

where \( y_{exp-i} \) is the \( i \)-experimental measurement of the dependant variable that corresponds to the normalized optical density measurement \( OD'_{600} \) at time \( t_i \) and \( y'(t_i, p) \) is the normalized value of the model. The variables were normalized taking into account that the sensor signals and the optical density have different orders of magnitude. Post-regression diagnostics are techniques that test the quality of the parameter estimation [34]. The methods determine the significance of the confidence intervals, the identifiability with cross correlations between the parameters, and the sensitivity of the output response under parameter perturbations [35]. Firstly, from the parameter estimation procedure of Equation 17 the confidence intervals of the parameters can be computed with

4.4. Model fitting and post-regression analysis

The parameter estimation method used here applies a minimization procedure [33]:

\[ \min_{p \in \mathbb{R}^n} E(p) \]  

(17)

where \( E(p) \) is the objective function that considers the difference between the simulation and experimental data at time \( t_i \) [33], with a set of parameters \( p \):
the inverse of the Fisher information Matrix (FIM) \cite{36}:
\[ C_J = \text{FIM}^{-1} = \frac{E(p)(J(p)^T V^{-1} J(p))^{-1}}{N - n_p} \] (19)

where \( J(p) \) is the Jacobian and the information on the standard deviation of the measurements \( (\sigma_i) \) is grouped in a diagonal matrix \( V \) \cite{33, 37, 38}. The parameters are practically identifiable if and only if the confidence intervals of the parameters are finite \cite{37, 38}. The Hessian of the quadratic error estimation \( (H = \nabla^2 E(p)) \) can also provide estimates of the confidence intervals \( (C_H = \frac{2}{N - n_p} E(p)H(p)^{-1}) \) \cite{33, 38, 39}, and the computation of the coefficient of variation is:
\[ \%\delta_{j,\text{variation}} = 100 \frac{\sqrt{C_{jj}}}{p_j} \] (20)

The confidence intervals using a two tailed t-student distribution \( (\delta_{j,\text{confidence}}) \) are \cite{36}:
\[ \delta_{j,\text{confidence}} = \pm t_{(N - n_p)\alpha/2} \sqrt{C_{jj}} \] (21)

where the \( C_{jj} \) is the diagonal element corresponds to the parameter in the matrix \( C_J \) or \( C_H \), \( \alpha \) is the confidence level, \( N \) is the number of experimental data points, \( n_p \) is the number of parameters and the difference \((N - n_p)\) are the degrees of freedom of the two tailed t-student distribution. Secondly, the cross-correlation between any two parameters can be evaluated with the correlation matrix \( (R_{ij}) \) \cite{40, 41}:
\[ R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \quad i \neq j, \]
\[ R_{ij} = 1 \quad i = j, \] (22)

where a value of \((\pm 1)\) indicates a complete dependency, and a value of zero
indicates no dependency between the parameters [41]. Thirdly, the sensitivity analysis examines the model response to parameter perturbation. The method involves differentiation of Equation [42, 43, 44]. In this study, we will consider the sensitivity with respect to the states:

\[ y_x = \frac{\partial g}{\partial x} \]  

(23)

For a linear model \((y = \beta \cdot \text{Signal} + \alpha)\), the sensitivity of the optical density with respect to the sensor signal equals to the coefficient \((\beta)\). Table 1 contains information on the sensitivity of the optical density measurements as a function of the sensor signals. Finally, the performance of each regression was evaluated with the mean square error (MSE), the root mean square error (RMSE), and the data estimated with the mean absolute percentage error (MAPE), expressed by the following equations:

\[ \text{MSE} = \frac{\sum(y_{\text{exp}} - y')(t_i)^2}{n_{\text{DOF}}} \]  

(24)

\[ \text{RMSE} = \sqrt{\text{MSE}} \]  

(25)

\[ \text{MAPE} = \frac{1}{N} \sum_i \frac{|y_{\text{exp}}(t_i) - y'(t_i, p)|}{y_{\text{exp}}(t_i)} \]  

(26)

where \(n_{\text{DOF}}\) is the number of degrees of freedom and \(N\) is the number of experimental data points for validation.

4.5. Regression

The biomass response was fitted with a linear regression equation using the normalized sensor signals \(\text{Signal}'\) data:

\[ OD'(t_i) = \beta \cdot \text{Signal}'(t_i) + \alpha \]  

(27)
where we assumed the data is an independent vector of time measurements. It is proposed that the signal output of the different sensors is described by an empirical model that follows a power law model, and is similar to a mass transfer expression \[45, 46, 47\], where the signal is a function of the optical density (OD\(_{600}\)), the stirrer speed (\(N\)), the superficial gas velocity (\(V_s\)), the stirrer diameter (\(T\)), the tank diameter (\(D\)), the number of blades (\(n_b\)), and the number of stirrers (\(n_s\)):

\[
\text{Signal} = K_1 \cdot (\text{OD}_{600})^b \cdot (N)^c \cdot (v_s)^d \cdot (T)^e \cdot (D)^f \cdot (n_b)^g \cdot (n_s)^h \tag{28}
\]

In our fed-batch experiments, the reactor configuration and the air flow did not change, then the equation becomes:

\[
\text{Signal} = K_2 \cdot (\text{OD}_{600})^b \cdot (N)^c \tag{29}
\]

where the constant \(K_2\) is a proportional constant, and the exponents \(b\) and \(c\) take into account the effect of each input on the signal output. The logarithmic operator in Equation 29 gives:

\[
\ln(\text{Signal}) = K + b \cdot \ln(\text{OD}_{600}) + c \cdot \ln(N) \tag{30}
\]

where \(K\) is a constant, and Equation 30 can be solved through multiple linear regression (MLR). The method consists of minimizing the difference between the experimental data and the model, and defining \(K\), \(b\) and \(c\). A 3D plot representation of the MLR result is presented in the graphical abstract of this paper.
4.6. Growth rate estimation

Growth rate estimation were computed with the time derivatives of the natural logarithm of the signal \([22]\). The optical density was assumed to be linearly correlated with the sensor signals:

\[ OD(t) - OD_0 = a \cdot \text{Signal}(t) - \text{Signal}_0 \]

Equation 27 is normalized and Equation 31 represents the non-scaled optical density. The initial signal value was set to zero \((\text{Signal}_0 = 0)\). Then, the time derivative of Equation 31 becomes:

\[ \frac{dOD}{dt} = a \cdot \frac{d\text{Signal}}{dt} \]

Using the approximation that the initial optical density tends to zero \((OD_0 \approx 0)\) (see Figure 2a), then Equation 8 becomes:

\[ \frac{1}{OD} \frac{dOD}{dt} = \frac{1}{\text{Signal}} \frac{d\text{Signal}}{dt} = \dot{\ln}(\text{Signal}) \]

Following this, from Equations 10 and 33 we get:

\[ \ln(OD_{600}) = \ln(\text{Signal}) = \mu - D \]

Equation 34 shows that the time derivative of the natural logarithm of the \textit{in-situ} signal is equal to the microbial growth rate \((\mu)\) and the dilution term \((D)\). For the sake of simplicity, we referred in this study to Equation 34 as growth rate.
4.7. Gaussian smoothing

We assumed that the signal outputs contain random noise (stems from the noise inherited in the measurements). The percentage error mean value of the signal output in steady state (at time interval 35-40 hr in our experiments) was computed using the following equation:

\[
\text{\% error} = \left| \frac{\text{Signal} - \bar{\text{Signal}}}{\text{Signal}} \right| \cdot 100\% \tag{35}
\]

where the bar represents the average value of the signal output during the steady state interval. The results of this percentage error are presented in Figure 7. A moving average smoothing technique of the signals was implemented in Matlab (Mathworks Inc.). In the computation, the signals are multiplied by a Gaussian-shaped weighting function:

\[
\text{Signal}'(t_j) = \frac{1}{\sum_{i=1}^{N} w_{ij}} \sum_{i=1}^{N} \text{Signal}(t_i) w_{ij} \tag{36}
\]

where the value of \( w_{ij} \) is

\[
w_{ij} = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(t_i-t_j)^2}{2\sigma^2}} \tag{37}
\]

with the indices \( i \) and \( j \) vary from 1 to \( N \), and the degree of smoothing is defined by the value of \( \sigma \) which is proportional to the full width at half maximum (FWHM) of the Gaussian function (FWHM \( \approx 2.35\sigma \) \[48\]). The objective of the Gaussian smoothing method is to improve the signal to noise ratio (S/N). In this study, the Gaussian filter was heuristically tuned, and the values of the signal window and their derivatives are in Table I.
4.8. Simulations and model fitting

The function `dividerand` in Matlab (Mathworks Inc.) was used to split the available datasets to training (70%) and validation (30%). The linear fittings and multiple linear regression algorithms were conducted with the functions `regress` and `fitlm`, the latter uses an algorithm based on Iteratively Reweighted Least Squares method. The Gaussian filter was implemented with the `smoothdata` function, and the derivatives with the `gradient` function. The window of the Gaussian filter was adjusted heuristically based on the input data, and the value of the window of the signal and its derivative is in Table 2.

5. Results

5.1. Growth experiment

Figure 2 presents the results of fed-batch *C. glutamicum* growth experiment in CGXII with an initial optical density of unity and an initial glucose concentration of 20 g L$^{-1}$. The results consist of the OD$_{600}$, measured offline by a spectrophotometer (Figure 2a), the glucose concentration, measured by an HPLC (Figure 2b), the total carbon dioxide excreted and the total oxygen consumed during the process (Figure 2c), the feeding strategy (Figure 2d), the dissolved oxygen in the medium (Figure 2e) and the stirring rate of the Rushton type impeller (Figure 2f). The OD$_{600}$ values increased during the batch phase until a steady-state was achieved when the initially provided sugar was depleted after 12 hours. A three hours substrate feeding started when the $pO_2$ was close to 100%, and the glucose inlet triggered a new exponential growth phase. When the substrate was consumed after three hours of feeding, the system stabilized on a new steady state at an OD$_{600}$ value of 36. The feeding strategy consisted of 0.1 L/hr for 1.5 hours, and 0.21 L/hr for 1.5 hours with a glucose concentration of 100 gL$^{-1}$. An overshoot was observed in the OD$_{600}$ measurements following
glucose depletion (time 25-30 hr), when the OD\textsubscript{600} transiently increased to 42 before it declined to a stable value of 36. The total CO\textsubscript{2} and O\textsubscript{2} concentrations increased during the batch phase until a steady state was achieved at hour 12, and following the feeding phase, the signals reached a new steady state after 30 hours (Figure 2d). The pO\textsubscript{2} was reduced during the batch phase, and when the pO\textsubscript{2} was below 30 \% the stirring rate was increased (programmed with a cascade controller) in order to elevate the oxygen mass transfer coefficient. Finally, when glucose was absent, the pO\textsubscript{2} concentration increased and the stirring controller settled on a value of 200 rpm. The dynamics of this process follows a typical fed-batch process [49, 50].

5.2. Quantitative and qualitative comparison of the sensors

In this study, growth of \textit{C. glutamicum} was monitored by offline measurements of the OD\textsubscript{600} using a spectrophotometer and the results were compared with signals of five different \textit{in-situ} probes. The sensors represent different measurement technologies, each with different sampling frequency and sensitivity (see Table 1 and Methods). The most frequent measurements were the ones integrated in Lucullus (Securecell, Switzerland) (NIR OPL05, NIR OPL01, and BlueInOne sensor), and the less frequent was the LS external. The numerical sensitivity of the OD\textsubscript{600} was the highest for the Optek OPL01 probe, implying that small perturbations in the sensor signal correspond to large changes in the OD\textsubscript{600}. The LS external sensor exhibited the lowest sensitivity, implying a robust OD\textsubscript{600} estimation.

Figure 3 presents normalized offline (OD\textsubscript{600}') measurements as a function of normalized signals (\textit{Signals}') obtained from the different probes. Each plot contains a confidence region of the linear regression as a gray area. Firstly, the intercept in all the sensors had a confidence interval in the order of the estimate, further acquisition of data in low optical density concentration can...
reduce the identifiability of the parameter in that region [51]. Secondly, the slopes of the regression curves of the sensors were close to one \((\alpha \approx 1)\), and the total CO\(_2\) slope (Figure 3f) was the most distant ideal slope, meaning that the deviation in the regression was the highest. This deviation emerged from the data points acquired during the time intervals 22-30 hr (see Figure 2a and 2c). At the beginning of each interval, a peak in the offline measurement was observed following sugar depletion, but no such peak was observed in the total CO\(_2\). This is likely due to the difference in growth and respiration following a sugar starvation period (see Discussion). Thirdly, the linear regressions of the sensor had relatively low RMSE values (RMSE \(\geq 0.15\)) whereas the MAPE was particularly high for the NIR OPL05. These observations are consistent with the non-linear behavior of the sensor data in Figure 2a. The external LS sensor had the closest regression to the ideal curve with lowest RMSE and MAPE, and thus exhibits the best linear regression performance. For all the sensors the cross correlation between the slope and the ordinate \((R_{\alpha/\beta})\) was in the range of 0.8-0.9, values that suggest a need of additional measurements at lower optical density values in order to improve identifiability of the ordinate.

A multiple linear regression algorithm was implemented between the sensors signal outputs as a function of optical density and stirring rate (Equation 30) using data from the fed-batch experiments. The results (Table 3) consist of a proportional constant \((K)\) between the signal output and the offline OD\(_{600}\) measurement, the exponents for the OD\(_{600}\) \((b)\), the stirring rate \((c)\), and the cross correlation of the parameters (Table 4).

5.3. Impact of Stirring and Air flow.

The results of the normalized signals as a function of the stirring rates ranging between 200-1100 rpm are presented in Figure 4. The effect of the stirring and aeration on the measurements was evaluated during the stationary phase
of the growth (following substrate depletion) in the bioreactor. Firstly, for the NIR Absorbance probes (Figures 4a and 4b) no significant variations in the signal below 500 rpm were observed. Above 500 rpm, the signal measurements were linearly increasing with the stirring rate by a constant standard deviation. The external LS probe displayed a linear increase with stirring values above 300 rpm, with a constant standard deviation. The BS reflectance probe displayed stable measurements below 500 rpm, but highly noisy signals above 500 rpm, with standard deviation which increases with the stirring rate (ranging between 0.003 to 0.05). Note that Figure 4e, corresponding to the capacitance sensor, has a different y-axis scale than the other figures. As the stirring varied, the signal perturbation exhibited a change of approximately 60% with a constant standard deviation. Finally, the total CO$_2$ signal was not significantly affected by the stirring rate, and had a relatively low standard deviation in the order of $10^{-3}$. Experiments with variations in the air flow during the steady state phase (Figure 5) indicated that the absorbance NIR probes were less affected by the air flow (Figure 5a and 5b). In contrast, the BS reflectance and the offline optical density measurements were more sensitive to variable air flow values (Figures 5b and 5d). The capacitance probe showed a lower dependency on the air flow than on the stirring rate (Figure 5d). Finally, the air flow had less than 2 % effect on the total CO$_2$ (Figure 5f). The maximum signal deviation was studied (Figure 6) with the air flow (gray bars) and the stirring rate (black bars). Firstly, it was observed that the absorbance NIR probes, the external LS, the offline spectrophotometer measurement and the total CO$_2$ were less sensitive to the inputs manipulation (less than 10 % deviation) compared to the capacitance and the BS reflectance probes. The BS reflectance probe had more than 20 % maximum deviation with the manipulation of the inputs. Finally, the capacitance probe was the most sensitive to changes in the stirring rate and the air flow.
5.4. Inferring growth by time derivatives

Growth in microbial culture can be approximated by differentiating the logarithm of the population size with time (Equation 34). The main advantage of this non-parametric estimation of growth is its rapid computation, available as soon as the measurement signals are acquired [52]. However, a steady state analysis of the measurements showed that the signals contained noise (Figure 7) (time interval 35-40 hr). The box plot diagram of the percent deviation from the mean during steady state (Figure 7) showed that the highest percent deviation was in the capacitance sensor (2.5 % percent deviation) and the NIR OPL05 sensor (1.8 % percent deviation), whereas the other sensors had lower than 1.5 % percent deviation. Generally speaking, all measurements contain some noise, partly due to human error and partly due to inherent limitations in the instruments, and the amplification of the noise in the signal by differentiation when computing the time signal derivative pose a major drawback for control applications. Noise can be reduced by accurate automatic sampling devices and by ‘soft sensors’, filtration algorithms that remove noise from signals [8]. In this study, the problem was addressed with a Gaussian filter implementation. The results of the biomass concentration and the time derivative of the logarithm (growth) are presented in Figure 8. The sensor signals tested (gray lines) followed the offline biomass concentration (black dots). There was discrepancy between the offline measurement and the in-situ absorbance NIR OPL05 signals (Figure 8a), and in the steady state of the total carbon dioxide measurements (Figure 8f). The offsets were consistent with the low RMSE values of these two sensors (Figure 3). The NIR OPL01 (Figure 8b), the LS external sensor (Figure 8c), the BS reflectance (Figure 8d), and the capacitance (Figure 8e) followed the process without an observed offset. Moreover, the capacitance sensor had a reduction in the signal during the feeding phase (time 22-25 hr). This reduc-
tion was not observed in the offline spectrophotometric measurements and the other sensor signals. The reason for this reduction was the increased stirring rate during the culture that modified the capacitance signal (Figure 4 and 6). Additionally, the total carbon dioxide (Figure 5) did not present a peak in its signal, in contrast to the offline measurement and the other sensors. Finally, the raw sensor data (gray lines) contained random noise that were amplified during the computation of the time derivatives, and thus reduced the quality of the estimation. A Gaussian filter implementation reduced the noise in the signals and their derivatives (blue lines).

6. Discussion

Understanding the cell culture conditions and the process dynamics requires a combined expertise of experimental methods and data analytics. Currently, sensor technologies with different hardware configurations and measurement principles are available in the market. The sensors are designed to estimate the biomass concentration by correlating their signals with the offline optical density measurements, cell dry weight and cell viability [53]. In this study, the in-situ biomass sensors signals were compared with the offline spectrophotometer measurement. Firstly, our analysis revealed that all the probes from this study were correlate to the growth (Figure 3). The $R^2$ values of the linear regressions were above 0.92, except for the NIR OPL05 sensor, in which the non-linearity seems to affect the linear regression results. The overshoot observed in the OD$_{600}$ measurements following glucose depletion (time 25-30 hr) is in agreement with previous studies that explored long term lactic acid adaptation [54], stress in the culture conditions [55], and cultures with mixture of glucose and maltose [56, 57]. We further dedicate a set of experiments to observe the linear regression between cell dry weight and optical density (See Appendix Figure A.10).
For the linear and multiple linear regression analysis, several error quantifiers were considered (Table 2): RMSE that evaluates the correlation accuracy, and MAPE that takes into account the accuracy in the testing dataset. In the linear regression, the RMSE values were below 0.15, and the highest MAPE was observed in the NIR OPL05.

Secondly, a multiple linear regression algorithm was implemented, as discussed in the Theory/Calculation section (See Equation 30). The algorithm describes the sensor signal outputs with a function that takes into account variations in the OD\textsubscript{600} values and the stirring rates (N) during the fed-batch cultivation (results of the regression are in Table 3). The small confidence intervals of the parameters $K$, and $b$ indicate a relatively precise estimation. However, the confidence intervals of the parameter $c$ were finite yet in some cases larger than the parameter itself, suggesting an inaccurate estimation. The closest to a linear regression with the factor $b \approx 1$, was the OPL01 probe, suggesting a linear relationship between the probe signals and the OD\textsubscript{600} values. The lowest effect of stirring on the probe signal was found in the NIR OPL05, with the exponential factor $c = -4.5 \cdot 10^{-3}$. In contrast, the highest effect of the stirring was observed in the capacitance sensor ($c = 0.11$). These results are in fact surprising because these NIR probes rely on absorbance of the near infra-red wave in the gap between the wave emitter and the probe’s detector, so larger gaps (e.g. OPL05) should absorb more noisy particles and cells as the stirring rate increases. Moreover, when evaluating the accuracy of the regression lines of the probes by the root mean square error (RMSE) and the regression coefficient, the LS external sensor was found to be the most accurate (RMSE=0.01, with $R^2$ values above 0.9.) However, the regression between the LS external signal and the OD\textsubscript{600} exhibited a highly non-linear behaviour ($b = 0.28$) and the values of the parameters $K$ and $c$ were highly cross correlated (Table 4), suggesting that
it is harder to separate the effect of the stirring rate from the bioreactor initial condition estimate. With other words, this analysis implies that the choice of biomass probe depends on the desired sensitivity and accuracy.

Thirdly, all the probes in this study were perturbed by the stirring rate and air flow (Figures 4, 5 and 6). The stirring rate had a higher impact on the signal compared to the air flow. The best performing technologies (the two NIR probes, the LS external sensor and the total CO₂ produced) under different conditions of stirring and air flow had a maximum deviation below 10%, and the percent deviation was similar to the 10% error of the off-line optical density measurement. Interestingly, the error in the offline measurements indicates that manual measurements do not guarantee noise free measurements. On the contrary, manual measurements can contain noise arising from inaccurate manual sampling and handling. The back scattering probe was sensitive to changes in the airflow, likely due to an increased interference with the air bubbles, perceived by the sensor as a signal (biomass). The observation that the airflow has a minor impact on the NIR probes is more surprising, because increased air bubbles (as the airflow is increased) are expected to be absorbed by the emitted NIR signals and thus interfere with the absorption caused by real microbial cells. A possible explanation suggests that increasing reactor airflow does not generate small enough droplets (1-5 mm sizes) that can absorb the near infrared waves in the NIR probe gap (between the emitter and the wave detector).

It is important to note that the capacitance was particularly sensitive to the stirring rate. In this study, there was a special interest in exploring the performance of the capacitance sensor due to its correlation with viability, cell geometry, and since membrane perturbation promotes the excretion of amino acids [58]. Furthermore, it is known that cell viability decreases during periods of substrate deprivation. This fact must be taken into account when monitoring
of growth and estimation of the microbial growth rate ($\mu$) in order to distinguish between viable and dead cells [59]. However, we found a strong bias in the signal (60 % bias) as the stirring rate changes, bias with a constant variance (Figure 4e). It can be argued that the capacitance sensor must be as far as possible from the reactor walls, baffles and other metal objects, as their presence close to the probe field causes signal interference [60]. This claim is in agreement with the noise observed in our 2.5 L lab-scale bioreactors, and we suggest to explore the capacitance technology in larger scale bioreactors.

The fact that the total CO$_2$ does not follow the peaks detected with the biomass sensors after glucose starvation can be an indication of a dynamical allocation of cellular resources. Previous studies in E. coli focusing on glucose starvation demonstrated a drift in the cellular metabolism (nutrients to precursors) and in the gene expression (precursors to proteins that produce biomass) [49, 61]. Furthermore, it has been reported that E. coli cells repress nutrient consumption for biomass production and use the remaining nutrient in the environment to delay cell death [62]. Monitoring and estimation of the parameters in microbial cell cultures is feasible with information of the oxygen uptake rate (OUR), the carbon excretion rate (CER) [13] and the amount of heat transfer [63]. However, cells may change their metabolic activities under different stress conditions, resulting in changing respiratory coefficient $RQ = \frac{\text{CO}_2 \text{-produced}}{\text{O}_2 \text{-consumed}}$. Thus, prolonged starvation periods in the fed-batch process can lead to different RQ profiles, that do not necessarily translate into a biomass growth signal. In this study, the total amount of CO$_2$ produced against the O$_2$ consumed (Figure 9) had a linear relationship, implying a constant slope. In bioprocesses, precise measurements of growth and growth rate have the potential to enhance control based applications.

Finally, the computation of the growth rate (Equation 34) amplified the
random noise contained in the signals (Figure 7). This problem was resolved using a Gaussian filter implementation with window sizes available in Table 2 and the non-parametric estimations of the growth rate were in the same order of magnitude as previous C. glutamicum reports ($\mu_m = 0.3 - 0.5h^{-1}$) [64].

7. Conclusions

The goal of this study was to evaluate the performance of different sensor technologies in a real-time Corynebacterium glutamicum monitoring. The sensors that were tested had different measurement principles, sensitivity and sampling frequency, including NIR absorbance, external light scattering, back scattering reflectance, dielectric spectroscopy (Capacitance) and off gas sensor measurements. The sensitivity of the sensors to air flow and stirring rates was evaluated, with the latter affecting all the sensors. The capacitance sensor was particularly sensitive to the stirring rate, and the effect was attributed to the relatively small bioreactor scales. Linear correlation and multiple linear regression between the different sensors and the offline optical density measurement enabled reliable online estimation of the growth in our bioreactor set-up. The studied sensors exhibited steady state noise that posed a drawback when computing the time derivatives. The problem was addressed with an implementation of a Gaussian filter, demonstrating that these technologies were also capable of providing an in-situ estimates of the growth rate.

8. Authors contribution

PL, NB and GS wrote the manuscript. PL and AT performed the experiments. NB supervised the project. All authors have read and approved the final submitted manuscript.
9. Acknowledgments

The authors would like to thank Haakon Eng Holck for support and discussion.

10. Grant Numbers

This research work has received funding from the Department of Chemical Engineering at NTNU, grant number 70441143, and was partially funded by the Novo Nordisk Fonden within the framework of the Fermentation-based Biomanufacturing Initiative (grant number NNF17SA0031362).
Bibliography


[14] D. Krämer, R. King, A hybrid approach for bioprocess state estimation


32


33
[42] R. Brun, M. Kühni, H. Siegrist, W. Gujer, P. Reichert, Practical identifi-
ability of ASM2d parameters—systematic selection and tuning of param-
S0043-1354(02)00104-5

[43] D. A. Tortorelli, P. Michaleris, Design sensitivity analysis: Overview and
1080/174159794088027573

[44] J. D. Stigter, J. Molenaar, A fast algorithm to assess local structural iden-
2015.05.004

[45] W. Klöckner, R. Gacem, T. Anderlei, N. Raven, S. Schillberg, C. Lat-
termann, J. Büchs, Correlation between mass transfer coefficient $k_{La}$ and
relevant operating parameters in cylindrical disposable shaken bioreactors
doi:10.1186/1754-1611-7-28

[46] F. Garcia-Ochoa, E. G. Castro, Estimation of oxygen mass transfer co-
efficient in stirred tank reactors using artificial neural networks., En-
s0141-0229(01)00297-6

[47] J. Villadsen, J. Nielsen, E. Keshavarz-Moore, Bioreaction Engineer-
978-1-4419-9688-6

[48] M. Feissel, W. Lewandowski, A comparative analysis of vondrak and gaus-
doi:10.1007/BF02523694


Figure 1: Scheme of the bioreactor set-up for the *C. glutamicum* experiments. The system has 1) pumps for acid, base, antifoam, and substrate 2) an inlet that allows the liquid inflow at the top of the bioreactor, 3) air valve, 4) air mass flow controller, 5) pH Controller (pHC) with a defined set-point (pH\(_{sp}\)) that triggers the acid/base pumps, 6) in-situ probe (green), 7) heating and cooling element, 8) pO\(_2\) controller with a defined set-point (pO\(_{2sp}\)) controlling the stirring of a Rushton type impeller, 9) temperature controller (TC) with a defined temperature set-point (T\(_{sp}\), 10) external biomass sensor, 11) multiplexer of the gas phase, 12) BlueInOne sensor that monitors O\(_2\) and CO\(_2\) concentrations in the gas phase. The liquid samples (green dashed line) are taken to the Numera. The Numera system consists of a 13) liquid multiplexer that takes sample from the different bioreactors, 14) a dilution module that can transfer the sample to a 15) filtration module, or to an 16) autosampler.
Figure 2: A fed-batch experiment presenting the growth as a function of time. Monitoring devices obtained information of a) biomass concentration from the optical density measurement, b) glucose concentration measured with HPLC, c) integral of the CO$_2$ produced and the O$_2$ consumed, d) feeding liquid inlet, e) oxygen dissolved with the pO$_2$ measurement, and f) Rushton-type stirrer inlet.
Figure 3: Linear regression of the normalized $\text{OD}_{600}'$ and the normalized sensor signal outputs. The experimental data determined with the spectrophotometer (black dots), the ideal slope (black solid line) and the linear correlations based on least square regression (black dashed lines). The gray areas are the 95% confidence interval of the regression. The correlations correspond to the normalized $\text{OD}_{600}'$ as a function of the normalized sensor signals of a) NIR OPL05, b) NIR OPL01, c) External LS, d) BS reflectance, e) Capacitance, and f) Total CO$_2$.
Figure 4: Normalized sensor signals as a function of the stirring. a) NIR OPL05, b) NIR OPL01, c) External LS, d) BS reflectance, e) Capacitance, and f) Total CO$_2$ produced.
Figure 5: The normalized sensor signals are affected by the stirring rate (abscissa) and airflow (ordinate). a) NIR OPL05, b) NIR OPL01, c) BS Reflectance, d) Capacitance, e) Spectrophotometer optical density measurement, f) Total CO₂.
Figure 6: The maximum percentage deviation with changing air flows (0.5-2 L, gray bars), and with changing stirring rates (200-1100 rpm, black bars). The data was obtained in the stationary phase of the process.
Figure 7: The sensors signals contain measurement noise. The box plot diagrams show the percentage error of the signal outputs with respect to their mean value (Equation 35) in steady state during the time interval 35-40 hr. The central mark in the box is the median, and the bottom and top edges indicate the 25th and 75th percentiles, respectively. The circles represent outliers, and the whiskers extend to the most extreme data points.
Figure 8: Normalized optical density and estimation of the growth rate in the bioreactor. The black dots are the offline OD'₆₀₀ measurement, the dashed black lines are the estimation of growth rate with the OD'₆₀₀, the gray areas are the 95% confidence intervals of the sensor signals with the linear regression model, the gray lines are the estimations of the derivative without a filter, and the blue lines are the signal with the application of a Gaussian filter.

a) NIR Optek OPL05, b) NIR OPL01, c) LS External, d) BS Reflectance, e) Capacitance, f) Total CO₂.
Figure 9: The total CO$_2$ produced and the total O$_2$ consumed in the culture exhibited a near-linear relationship.
Table 1: Sensors, measurement principle, sample frequency, sensitivity, signal window $(w_1)$, and window of the derivative $(w_2)$

<table>
<thead>
<tr>
<th>Sensors</th>
<th>Measurement principle</th>
<th>Frequency [hz]</th>
<th>Sensitivity [OD$_{600}$ Signal$^{-1}$]</th>
<th>$w_1$</th>
<th>$w_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optek NIR OPL05</td>
<td>Near Infrared absorbance</td>
<td>$1 \cdot 10^{-1}$</td>
<td>22</td>
<td>2581</td>
<td>2950</td>
</tr>
<tr>
<td>Optek NIR OPL01</td>
<td>Near Infrared absorbance</td>
<td>$1 \cdot 10^{-1}$</td>
<td>112</td>
<td>3171</td>
<td>1649</td>
</tr>
<tr>
<td>LS External Acquilab</td>
<td>External light scattered source</td>
<td>$5 \cdot 10^{-4}$</td>
<td>0.05</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Buglab proBE 3000</td>
<td>Back-scattering (BS) reflectance</td>
<td>$1.6 \cdot 10^{-2}$</td>
<td>0.29</td>
<td>996</td>
<td>623</td>
</tr>
<tr>
<td>Aber Futura Capacitance</td>
<td>Dielectric spectroscopy</td>
<td>$3.3 \cdot 10^{-2}$</td>
<td>5.1</td>
<td>1011</td>
<td>632</td>
</tr>
<tr>
<td>BlueInOne Ferm sensor</td>
<td>Infrared (IR) (CO$_2$, ZrO$_2$ (O$_2$))</td>
<td>$1 \cdot 10^{-1}$</td>
<td>1.8</td>
<td>2581</td>
<td>2950</td>
</tr>
</tbody>
</table>
Table 2: Linear regression results of the normalized OD'$_{600}$ as a function of the normalized sensor signal (OD'$_{600}$ = $\alpha \cdot $Signal' + $\beta$), R-squared ($R^2$), root mean square error (RMSE), cross-correlation between the slope and the ordinate ($R_{\alpha/\beta}$) and mean average percentage error of the predictions (MAPE). The $p$-values were lower than 0.05, indicating a linear relationship between the response and the predictor variables.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>$R_{\alpha/\beta}$</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR OPL05</td>
<td>$-0.12 \pm 0.16$</td>
<td>$0.9 \pm 0.2$</td>
<td>0.87</td>
<td>0.14</td>
<td>0.91</td>
<td>0.19</td>
</tr>
<tr>
<td>NIR OPL01</td>
<td>$-0.14 \pm 0.09$</td>
<td>$1 \pm 0.1$</td>
<td>0.95</td>
<td>0.14</td>
<td>0.91</td>
<td>0.13</td>
</tr>
<tr>
<td>LS External</td>
<td>$-0.1 \pm 0.17$</td>
<td>$1 \pm 0.1$</td>
<td>0.94</td>
<td>0.08</td>
<td>0.9</td>
<td>0.06</td>
</tr>
<tr>
<td>BS Reflectance</td>
<td>$-0.02 \pm 0.1$</td>
<td>$0.9 \pm 0.14$</td>
<td>0.96</td>
<td>0.07</td>
<td>0.84</td>
<td>0.15</td>
</tr>
<tr>
<td>BS Capacitance</td>
<td>$-0.1 \pm 0.17$</td>
<td>$1 \pm 0.2$</td>
<td>0.92</td>
<td>0.09</td>
<td>0.92</td>
<td>0.12</td>
</tr>
<tr>
<td>Total CO$_2$</td>
<td>$0.1 \pm 0.1$</td>
<td>$0.8 \pm 0.17$</td>
<td>0.92</td>
<td>0.08</td>
<td>0.81</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Table 3: Multiple linear regression analysis of the signals as a function of OD$_{600}$ and the stirring rate ($\ln(\text{Signal}) = K + b \cdot \ln(\text{OD}_{600}) + c \cdot \ln(N)$). The p-values of all the correlations were lower than 0.05, indicating a significant multiple linear regression relationship between the response and the predictor variables.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>$K$</th>
<th>$b$</th>
<th>$c$</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR OPL05</td>
<td>$-1.07 \pm 0.28$</td>
<td>$0.51 \pm 0.05$</td>
<td>$0.004 \pm 0.04$</td>
<td>0.87</td>
<td>0.06</td>
<td>0.09</td>
</tr>
<tr>
<td>NIR OPL01</td>
<td>$-4.0 \pm 0.5$</td>
<td>$1.07 \pm 0.08$</td>
<td>$0.05 \pm 0.07$</td>
<td>0.94</td>
<td>0.14</td>
<td>0.73</td>
</tr>
<tr>
<td>LS External</td>
<td>$5.9 \pm 0.1$</td>
<td>$0.28 \pm 0.01$</td>
<td>$0.03 \pm 0.01$</td>
<td>0.99</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>BS Reflectance</td>
<td>$-1.5 \pm 0.8$</td>
<td>$1.3 \pm 0.15$</td>
<td>$0.07 \pm 0.1$</td>
<td>0.9</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>Capacitance</td>
<td>$-3.6 \pm 1.4$</td>
<td>$1 \pm 0.23$</td>
<td>$0.11 \pm 0.18$</td>
<td>0.9</td>
<td>0.33</td>
<td>0.92</td>
</tr>
<tr>
<td>Total CO$_2$</td>
<td>$-6.9 \pm 1.6$</td>
<td>$1.5 \pm 0.2$</td>
<td>$-0.06 \pm 0.26$</td>
<td>0.97</td>
<td>0.14</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Table 4: Cross correlation of the parameters obtained from the multiple linear regression algorithm \((\ln(\text{Signal}) = K + b \cdot \ln(O_{D600}) + c \cdot \ln(N))\)

<table>
<thead>
<tr>
<th>Sensor</th>
<th>(R_{Kb})</th>
<th>(R_{Ke})</th>
<th>(R_{bc})</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR OPL 05</td>
<td>-0.58</td>
<td>-0.76</td>
<td>-0.11</td>
</tr>
<tr>
<td>NIR OPL 01</td>
<td>-0.37</td>
<td>-0.86</td>
<td>-0.13</td>
</tr>
<tr>
<td>LS External</td>
<td>-0.43</td>
<td>-0.94</td>
<td>0.13</td>
</tr>
<tr>
<td>BS Reflectance</td>
<td>-0.64</td>
<td>-0.79</td>
<td>0.05</td>
</tr>
<tr>
<td>Capacitance</td>
<td>-0.53</td>
<td>-0.84</td>
<td>0.00</td>
</tr>
<tr>
<td>Total CO(_2)</td>
<td>-0.25</td>
<td>-0.91</td>
<td>-0.12</td>
</tr>
</tbody>
</table>
Appendix A.

Appendix A.1. Cell dry weight and Optical density

During the growth experiments, the cell dry weight and the optical density had a linear relationship (Figure A.10). The data was normalized since each spectrophotometer will change the value of the slope.

![Graph showing the linear relationship between cell dry weight and optical density](image)

Figure A.10: The normalized cell dry weight (CDW') of *C. glutamicum* exhibited a linear relationship with the normalized OD'_{600} measurement. The black dots are the experimental measurements, the black line represents the linear model with a unitless slope of 1.03 and a confidence interval of [0.96, 1.1]. The gray area represents the 95% confidence interval of the regression. *p*-value $= 1.7 \cdot 10^{-23}$, RMSE $= 0.072$.

Appendix A.2. Filter implementation

Figure A.11 presents the OD$_{600}$ and the NIR results from growth experiments. The noise in the sensor measurement (gray line) is amplified during the computation of the derivatives, suggesting the need to use a filter in order to reduce the noise. A Savitzky-Golay filter (black lines) has an overshoot in the beginning of the culture, whereas a Gaussian filter does not have an overshoot, and it is smooth at the beginning of the bioprocess.
Figure A.11: Growth experiment and estimation of growth rate measuring optical density and a NIR probe. The black dots correspond to the data obtained from the offline OD$_{600}$ measurements, the gray lines are the NIR sensor signals without a filter, the black dashed lines are the signals with a Gaussian filter and the continuous black line is the signal with a Savitzky-Golay filter. a) Optical density, b) NIR measurement, c) Estimation of microbial growth rate.
Appendix A.3. Stirring and Air Flow experiments

Perturbations of the Stirring and Air flow were programmed in Lucullus (Securecell, Switzerland). Figure A.12 presents an example of the NIR signal perturbation with different values of Stirring and Air flow.

Figure A.12: Perturbations of the NIR signal with different values of air flow and stirring rates a) NIR measurement b) \( pO_2 \) concentration, c) Stirrer and d) Air flow.