Fast measurement of phosphates and ammonium in fermentation-like media: a feasibility study

Pontius, Katrin; Praticò, Giulia; Larsen, Flemming H; Skov, Thomas; Arneborg, Nils; Eliasson Lantz, Anna; Bevilacqua, Marta

Published in:
New Biotechnology

Link to article, DOI:
10.1016/j.nbt.2019.11.006

Publication date:
2020

Document Version
Peer reviewed version

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Fast measurement of phosphates and ammonium in fermentation-like media: a feasibility study

Katrin Pontius, Giulia Praticò, Flemming H. Larsen, Thomas Skov, Nils Arneborg, Anna Eliasson Lantz, Marta Bevilacqua

PII: S1871-6784(19)30041-X
DOI: https://doi.org/10.1016/j.nbt.2019.11.006
Reference: NBT 1223
To appear in: New BIOTECHNOLOGY

Received Date: 5 February 2019
Revised Date: 7 November 2019
Accepted Date: 18 November 2019


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier.
Fast measurement of phosphates and ammonium in fermentation-like media: a feasibility study

Katrin Pontius\(^a\), Giulia Praticò\(^{b,c}\), Flemming H. Larsen\(^b\), Thomas Skov\(^b\), Nils Arneborg\(^b\), Anna Eliasson Lantz\(^a\), Marta Bevilacqua\(^{ab}\) marta@food.ku.dk

\(^a\) Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 229, 2800 Lyngby

\(^b\) Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg C, Denmark.

\(^c\) Department of Nutrition, Exercise and Sports, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg C, Denmark.

\(^\star\) CORRESPONDING AUTHOR: Marta Bevilacqua, Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg C, Denmark.

e-mail: phone: +45 52729860

HIGHLIGHTS

- Real-time monitoring of fermentation parameters is important for modern industries.
- Analysis of phosphates and ammonium in a wide range of concentration (10-100 mM).
• Use of synthetic samples provided a dataset with completely uncorrelated analytes.
• Mid and Near infrared spectroscopy were evaluated in fermentation-like media.
• Chemometric models on Mid-infrared spectra had prediction errors lower than 5%.

ABSTRACT

Real-time monitoring of bioprocesses plays a key-role in modern industries, providing new information on full-scale production, thus enabling control of the process and allowing it to run at optimal conditions while minimizing waste. Monitoring of phosphates and ammonium in fermentation processes has a twofold interest: they are important nutrients for living organisms while at the same time constituting environmental nutrient pollutants, for which unnecessary use and disposal must be avoided. In this report, the possibility of simultaneous analysis of phosphates and ammonium in fermentations was verified using spectroscopy-based methods combined with chemometrics to construct calibration models. To achieve this, the models were based on synthetic samples mimicking real fermentation media, providing a dataset where the analytes were completely uncorrelated. Different at-line techniques (mid- and near- infrared spectroscopy, MIR and NIR) were evaluated for their ability to monitor quickly both analytes, in a wide range of concentrations (10-100 mM), in three media of different complexities. Partial Least Squares (PLS) models on MIR spectroscopy gave very good results, with prediction errors lower than 5% for both analytes in all datasets. In contrast, the results for PLS models on NIR spectroscopy were inferior (prediction errors between 3 and 26%) for both analytes, as, in the case of phosphate, it could be demonstrated that the model was based on indirect predictions.
**KEYWORDS**

Process monitoring; near-infrared spectroscopy; mid-infrared spectroscopy; Chemometrics; Fermentation.

**ABBREVIATIONS**

ATR: attenuated total reflectance; CV: Cross-Validation; d1: recycle delay; DTGS: deuterated triglycerine sulfate detector; FT: Fourier Transformed; $h_i$: leverage of the $i$th validation sample; $i$: samples index; MIR: mid-infrared spectroscopy; MSC: Multiplicative Scatter Correction; MSEC: mean squared error in calibration; $N$: number of samples; NET: net analyte signal; NIR: near-infrared spectroscopy; NMR: Nuclear Magnetic Resonance spectroscopy; P: inorganic phosphates; PLS: Partial Least Squares regression; $R^2$: coefficient of determination; RMSE: Root Mean Squared Error; SiC: Silicon Carbide; SNV: Standard Normal Variate; $s_{PE}^2$: prediction uncertainty (variance); $s_{PEi}$: standard deviation; UV-VIS: Ultra-Violet Visible spectroscopy; VIP: Variable Importance in Prediction.

**INTRODUCTION**

Due to the constant need for more efficient and sustainable biotechnological production achieved by reducing consumption of energy and raw materials while improving yields, process understanding has particular importance. Real-time monitoring of bioprocesses plays a dual key-role, allowing industry to gain new knowledge on full-scale production and hence to be more prudent in the use of resources, while also enabling control over the process, with the possibility of running it under more
optimal conditions, minimizing waste. Both aspects hold not only for industrial production, but are equally important for efficient fermentation optimization.

The monitoring of phosphates and ammonium in fermentation processes is of particular interest. As important nutrients of living organisms, their measurement is crucial to avoid a deficiency in the fermentation broth. At the same time, they constitute nutrient pollutants and their unrestricted disposal affects water resources worldwide, compromising ecosystems and human health. In addition, phosphorus availability is finite and its unnecessary use and disposal must be carefully considered by any industry [1,2]. Consequently, for sustainable and economically viable production, excess addition of phosphates and ammonia to fermentation media should be avoided.

One of the most frequently used methods to quantify analytes in fermentation broth is colorimetric, based on chemical or enzymatic reactions that act selectively on the compounds to be measured, transforming them into species that can be observed and quantified by UV-VIS spectrophotometry. This analysis is generally applicable only in limited concentration ranges and involves a pretreatment of the sample before analysis: samples must be removed from the bioreactor, filtered, diluted, and conditioned before analysis [3]. The main drawback of this approach is the time delay in the process data between sampling and analytical result. On average, colorimetric assays take at least 30 min before a result is available. Moreover, for practical reasons such as equipment scheduling and operational shifts, samples taken during the process are usually collected and analyzed after the process has completed. Thus, results from assays describing process progress and performance are commonly available only after the fermentation has finished. In contrast, real-time techniques have the advantage of providing a direct insight into the bioprocess that is essential for early detection of problems, allowing immediate action without curtailing the process [4,5]. Nevertheless, real-time methods are still quite rarely applied to industrial scale bioprocesses for practical reasons: the analysis must preserve the sterility of the process, while at the same time being able to measure specific
analytes, sometimes at very low concentrations, in the presence of complex nutritive media [5]. Different strategies have been proposed for on-line monitoring of phosphates and ammonium during fermentation and elsewhere [3,6-10]. The most promising techniques implemented in on-line or at-line set up are vibrational spectroscopies (namely near-infrared and mid-infrared spectroscopy, NIR and MIR), coupled with mathematical modelling to relate the spectra to the process parameters to be measured [4,11-18]. Spectroscopic techniques have the advantages of enabling rapid analysis (< 3 min per spectrum depending on the instrument specifications and settings, and in the best case only about 10-20 sec) and can be applied on-line to the bioreactors by means of fiber-optic probes. In addition, multiple analytes can be quantified based on a single spectrum. When analyzing fermentation samples, however, mathematical models are generally challenged by the presence of numerous different components in the broth absorbing in the same spectral regions and quite often directly or indirectly correlated with each other. The resulting spectra have a high number of broadly overlapping features changing throughout the course of the fermentation. Multivariate calibration models will therefore often be valid only for processes conducted under the same conditions as that used to build them, and will fail if any slight unforeseen change occurs in the future. For meaningful and robust calibration models, the analyte of interest in the calibration samples should not be correlated with other species in the medium [19], which is rarely the case in real-life fermentation processes. Thus, there is a need for decoupling (or un-correlating) the concentration dynamics of the target analytes from those of other compounds present, introducing some artificial, uncorrelated variations but keeping the sample matrix as representative of the original fermentation medium as possible.

Although it has been argued that it is impossible to ignore the correlations/interactions among media components [15,20-22], others advocate the need to overcome the natural correlations and have addressed this challenge in different ways that fall into two main approaches [11,16-18,23-28]. The
first is the preparation of synthetic samples (alone or in combination with real samples in the model) [16-18, 23] while the second is an intermediate situation where samples are spiked with different analytes to break the correlations (so called, semi-synthetic samples or adapted samples) [24-28].

The objective here was to verify the possibility of simultaneous analysis of phosphates and ammonium in fermentation samples using spectroscopy-based methods. The calibration models were built on synthetic samples that were generated to resemble real fermentation media and at the same time provide a dataset with completely uncorrelated analytes. Although the determination of ammonia by both MIR and NIR spectroscopy has been described previously [4,11], a technique able to simultaneously quantify phosphates and ammonium in fermentation broth and in such wide ranges of concentrations (10-100 mM) has not been addressed so far. NIR and MIR spectroscopy were evaluated and compared for the ability to determine the concentration of these analytes. NIR has the advantage of being the easier to implement on-line on an industrial-scale fermenter, whereas MIR can be implemented on-line but requires shorter optical fibers and therefore a closer position to the reactor. The mid-range of IR is generally more sensitive and informative than NIR, providing a better likelihood of simultaneous detection of the two analytes in the lower concentration range.

**MATERIALS AND METHODS**

A more complete and detailed explanation is available in Supplementary material.

**Sample preparation**

To study efficiently the possibility of determining phosphates and ammonia by fast spectroscopic techniques, three different sets of synthetic samples were prepared by mixing different amounts of two bulk solutions of two salts containing the analytes of interest (0.4M Na$_2$HPO$_4$, and 4M NH$_4$Cl) in pure milliQ water, minimal medium and complex medium, respectively. For all three sets, 25
solutions were prepared containing both analytes in concentrations ranging from 10 – 100 mM equally spaced according to a full factorial design (Figure 1 and Table S1). The free- ware Visual MINTEQ 3.1 [29] was used to investigate precipitation (attributed to struvite (NH₄MgPO₄·6H₂O) and hydroxyapatite (Ca₅(PO₄)₃(OH)) [30, 31]) in the minimal medium during some initial preparation tests. A pH value of 6.0 was chosen for all sample sets to avoid such precipitation. This is also a widely used pH value in fermentation processes.

The minimal medium used for the second set of samples was prepared as described previously [32]. The complex medium in the third set of samples was prepared similarly, with addition of a complex source of nitrogen (potato protein, 55 g/l), which was added to the medium and enzymatically hydrolyzed using 2.75 g Alcalase (Alcalase® 2.4 L FG (EC 3.4.21.62), Unit Activity 2.4 AU-A/g, 105 Novozymes, Bagsvaerd, Denmark) for 2 h at 55°C.

Reference measurements

Phosphate Colorimetric Assay. Phosphate concentrations were determined in duplicate using the Phosphate Colorimetric Assay Kit (MAK030, Sigma Aldrich, St. Louis, USA), conducted on a 96 well plate (Costar Assay Plates, REF. 3370) and the Multiskan Go spectrophotometer (Thermo Scientific, Ratastie, Finland).

Ammonium Colorimetric Assay kit. Ammonium concentrations were determined in duplicate using the Ammonium Cuvette -Test 47-130 mg/L NH₄-N (LCK 302) (Hach Lange, Manchester, UK) and the Hach Lange Spectrophotometer DR 3900.

³¹P NMR spectroscopy. For experiments in water and minimal media, samples for ³¹P NMR spectroscopy were prepared in 5 mm (o.d.) NMR sample tubes by mixing 495 µl sample with 55 µl D₂O. For the samples from the complex media and the fermentation, 440 µl samples were mixed with 110 µl D₂O. All samples were analyzed using a Bruker (Rheinstetten, Germany) Avance DRX 500
(11.7 T) spectrometer operating at a Larmor frequency of 202.4 MHz for $^{31}$P, using a single pulse (90 degrees flip angle) experiment. Spectra were processed by using Bruker TopSpin 3.5 (Bruker BioSpin GmbH, Germany) and data were imported in MatLab 8.3.0.532 (MathWorks, Natick, MA).

Spectroscopic measurements

**Mid-Infrared spectroscopy (MIR).** Infrared spectra of all solutions were acquired in duplicate on a FT-spectrometer (MB100, ABB Bomem, ABB Bomem Inc., Québec, Canada) in attenuated total reflectance (ATR) mode, by putting 800 μL of each sample on the ATR cell (ZnSe crystal, 45°C) and recording the spectrum with 64 scans at a resolution of 4 cm$^{-1}$. PLS models were built on the whole spectral region (800 – 2250 cm$^{-1}$) as well as on the region related to the target species (950 – 1200 cm$^{-1}$ for phosphates and 1360 – 1540 cm$^{-1}$ for ammonium).

**Near Infrared spectroscopy (NIR).** Near infrared spectra were acquired in triplicate using a dispersive/scanning-rating NIR instrument (Foss DS2500, Foss Analytical A/S, Hilleroed, Denmark) in a slurry cup sample holder, using a golden reflector allowing an optical path of 1 mm (Foss DS2500, Foss Analytical A/S, Hilleroed, Denmark). PLS models were built on both the whole NIR spectral region (400 and 2500 nm) and on selected interval related to phosphates (1600-1900 nm) and ammonium (2050-2300 nm).

Data analysis

All data were imported by MatLab 8.3.0.532 and any further elaboration was performed using both in-house scripts and functions, and the PLS toolbox 7.8 (Eigenvector Research, Inc., Manson, WA). Among the whole set of 25 samples for each of the matrices, 8 were selected as a validation set (red color in Figure 1 and Table S1 and also circled in Figure 1). Considering the small overall number of samples for each data set, the validation samples were chosen to span the experimental space as far as possible without compromising the quality of the calibration set.
Spectra were preprocessed to eliminate unwanted physical variability by comparing the PLS results obtained based on un-preprocessed spectra with the most common and effective spectral preprocessing methods (Multiplicative Scatter Correction [42,43, 51], Standard Normal Variate [51], first and second derivative [51]). The method leading to the optimal PLS performances was chosen for each data set independently.

The Partial Least Squares (PLS) models [33-35] were assessed and compared for predictive performances mainly based on their Prediction Error in % computed according to the statistical calculations reported in [36]. Other parameters used to evaluate the PLS models and mentioned in the Figures are typical indicators of a PLS model and a short explanation of their meaning can be found in the Supplementary material.

RESULTS AND DISCUSSION

The three sets of samples prepared in pure water, minimal medium and complex medium, were analyzed by two different vibrational spectroscopic techniques (MIR and NIR) using colorimetric assays as reference methods for determination of phosphates and ammonium content. The absolute quantification of phosphates was also obtained by means of $^{31}$P NMR spectroscopy. This approach is highly selective as only species containing phosphorous are detected. Furthermore, only a simple sample preparation (addition of D$_2$O) is required, which ensures a higher reliability of this method as a reference analysis. An accurate analysis and comparison of all the reference methods was performed and reported in Supplementary Appendix A.

Given the higher uncertainty (Supplementary Figure A1, standard deviation values) in the reference values of phosphates obtained by the colorimetric kit than that obtained by NMR (< 2.4 mM), and the better agreement of the NMR results with theoretical values, the following regression models for phosphates were built on the NMR results. The ammonium models were built on the values obtained
by the colorimetric kit that were considered satisfactory. The choice of reference method is as important as the calibration process itself, as the model can never be better than the reference method.

Infrared Spectroscopy

The MIR spectra were investigated first for the potential to analyze phosphates and ammonium simultaneously as this type of spectrum is in general more informative than NIR.

Raw spectra inspection and interpretation. The raw IR spectra of the three sets of solutions, shown in Figure 2, were inspected to verify if quantification of the salts under study was possible and to identify spectral regions where they had signals. Once the spectra were colored according to the analyte concentration, two specific spectral regions were identified as concentration dependent (Figure 2B, C, E, F, H, I). The intense peak centered at 1636 cm$^{-1}$ dominating all spectra was due to water that is the main component in all the solutions (Figure 2A, D, G). The P-O stretching region (800-1400 cm$^{-1}$ [37]) displayed several peaks, which systematically changed intensity according to the concentration of phosphates in the solution for all three different sets of synthetic samples (Figure 2, inserts B, E and H). These are due to the symmetric and antisymmetric stretching modes of the P-OH and P-O$_2$ bonds in HPO$_4^{2-}$ and H$_2$PO$_4^-$, respectively, and their complexes [37-39]. The shape and the position of these peaks varied among the three media due to the differences in composition of the solutions. Nevertheless, the dependence of absorbance in this region on the concentration of phosphates in the samples was evident. Figure 2C, F and I show a peak with maximum intensity around 1455 cm$^{-1}$, assigned to the triply degenerate bending mode of the tetrahedral NH$_4^+$ ion [40,41]. Its dependence on the concentration of salt added was evident for all three sets of samples, despite the low intensity.

PLS Modeling. To determine the concentration of the two compounds in the samples, the data were then analyzed by PLS [33-35]. For all three sets of synthetic samples two different PLS1 models were
built for phosphates and ammonium concentrations. The spectra were pre-processed by Multiplicative Scatter Correction (MSC) [42,43] and mean-centering, and different models were built on the whole spectral region (800 – 2250 cm\(^{-1}\)) as well as on the region related to the target species (950 – 1200 cm\(^{-1}\) for phosphates and 1360 – 1540 cm\(^{-1}\) for ammonium). For clarity, the results are described focusing on the data set describing the complex medium samples (Figure 3), as they have the closest resemblance to the fermentation matrix. The results of the other data sets, describing the water and minimal medium samples, can be regarded as similar, see Supplementary Material, Table S4 and Figure 4.

The results for the entire spectral region (800 – 2250 cm\(^{-1}\)) are shown in Figure 3A, B, C for the prediction of phosphate concentrations and Figure 3D, E, F for ammonium. Both models were successful using 3 and 4 components, resulting in a Root Mean Square Error of Prediction (RMSEP) of 5.6 and 2.5 mM, respectively (Figure 3A, D), corresponding to a prediction error of 3.63\% and 3.15\% (Figure 4A, B).

Inspection of the Variable Importance in Prediction (VIP) index [44] (Figure 3B, E) and regression coefficients (Figure 3C, F) for the models of each species strongly highlighted one specific peak as the most important in building the model (selected bands in Figure 3). The two selected bands are coherent with the respective regions identified and known in literature [37–41] to be correlated with specific vibrational modes of the phosphates and ammonium ions, namely peaks at 1076 cm\(^{-1}\) for phosphates and 1454 cm\(^{-1}\) for ammonium. This finding confirms that the model is built on direct correlations with relevant wavelength and can be considered trustworthy and robust. Focusing only on the interval related to phosphates and ammonium, highlighted in Figure 3, new PLS models were computed on these selected intervals. A comparison between the models built on the full region (800 – 2250 cm\(^{-1}\)) and the region related to the target species (950 – 1200 cm\(^{-1}\) for phosphates -highlighted bands in Figure 3B, C- and 1360 – 1540 cm\(^{-1}\) for ammonium -highlighted bands in Figure 3E, F) for
their most important parameters is shown in Figure 4A, B (and in Supplementary Table S4). In Figure 4 the results of the prediction errors for all models (full and selected region for phosphates and ammonium, both techniques) are summarized and compared.

The first and most general consideration to be drawn from Figure 4A, B was that overall the errors in the models increased when passing from water to minimal and complex medium due to increased complexity of the matrix. This behavior was expected due to the higher number of interfering species present in the complex matrix, that can hamper the model or that might react with the analytes in different ways altering their spectral signals or their response to the kits.

The VIP indices in Figure 3, showed only a very narrow region of the spectra of very high values, while the remaining parts of the spectra had VIP indices <1. This originally led to the hypothesis that models constructed only on these regions would not change or would slightly increase the performances of the PLS models. This was found to be partially correct, according to Figure 4 and Supplementary Table S4 as many of the models computed on the selected region showed quite similar, but slightly poorer performance than those on the whole region. This suggests that even when the VIP index of the major part of the spectra is below or very close to 1, the presence of these areas in the model contributes to its stabilization, especially improving selectivity towards the target species. The prediction error was investigated in more detail and the result of the analysis (Appendix B) confirmed the proposed hypothesis, showing that the intervals (for both analytes) were less sensitive than the whole region.

Some attempts to quantify ammonium by MIR spectroscopy have been published [4,12,14], however predominately, it has been reported by NIR spectroscopic methods [5,9,11,16,20,21,24,26,48]. Reports on quantification of phosphates in fermentation media by vibrational spectroscopy are in contrast quite limited [12-14], and could only focus on the use of MIR, since there are no specific
signals originating from phosphates in the NIR region. While the performance of ammonium quantification reported here with MIR spectroscopy align with some of the best results obtained elsewhere [4,12-14] (between 2 and 5% prediction error), those obtained in previous studies for quantification of phosphates are generally poorer than those ones obtained here [13-14] (about 10% prediction error) and only in rare cases similar [12] (<4% prediction error). The regions selected in literature upon which the models were built [4,12-14] match the regions selected in this work, for both analytes (for phosphates the peaks around 950 and 1100 cm\(^{-1}\) and for ammonium the region below 1500 cm\(^{-1}\)).

Near Infrared Spectroscopy

Although providing less detailed information than the MIR spectra, a NIR on-line monitoring probe would be the easiest measurement to implement for a fermentation process and thus, this technique was also investigated here.

**Raw spectra inspection and interpretation.** As for the MIR spectra, the raw NIR spectra of the three sets of solutions were visually inspected by overlaying to verify if quantification of phosphates and ammonium was possible. The identification of any spectral region that varied according to the content of these analytes was not easy to detect, so the correlation coefficient of each spectral wavelength with the reference values of each analyte was computed (Supplementary Figure S1). The spectral regions with the highest correlation with concentration of the analytes were the same in the three batches of samples. They correspond to bands reported for the two compounds under study [45-47]. Those correlated with phosphates were attributed to the third overtones of the P–O bonds in phosphate salts, reported between 1600 and 2000 nm [45]. In general, phosphates and other inorganic P compounds are hardly detectable by NIR spectroscopy due to the weak P–O dipole moment [46, 47]. Consequently, the correlation was less prominent for phosphates, with a correlation factor <0.7.
(Supplementary Figure S1), while about 0.8 for ammonium. Bands between 2100 and 2400 nm (attributed to the second overtones of the N-H bonds in NH$_4^+$ [45]) have been used previously for determination of ammonium in samples similar to those under study [16, 48, 49] and other examples are listed in [11]. However, the main purpose of the correlation analysis was to verify if any spectral region could be identified easily as respondent of phosphates or ammonium, confirming the attribution of the peaks hypothesized based on the literature [11,16,45,47-49].

**PLS modeling.** All spectra were pre-processed by mean-centering, which was chosen over Standard Normal Variate (SNV) and first and second derivative because it led to better results (data not shown). For each set of samples, two different PLS models were built on the whole NIR spectral region for phosphates and ammonium, respectively.

Inspection of the results showed that the PLS models based on the NIR data of minimal and complex media for both analytes allowed a good quantification of the compounds with errors in prediction between 3.1 and 7.1% (Supplementary Table S5). As an example, Figure 5 presents the models for phosphate and ammonium in complex medium, and a summary of the performance of all models is shown in Figure 4, next to the results obtained with IR. The VIP index and the regression coefficient of the models indicated the spectral regions 1600 – 2000 nm for phosphates (Figure 5B, C) and 2100-2400 nm for ammonium (Figure 5E, F) as particularly important for building the respective models. This improved the understanding of the predictions: while for ammonium, a true correlation between the most important variables (wavelengths) and the analyte was found (VIP scores and regression coefficients high and positive within the expected important spectral region, highlighted in blue in Figure 5E, F), the phosphates prediction strongly suggested that the model is based on an inverse correlation with the amount of water and is therefore not truly related to the concentration of phosphate. This can be concluded from the region around 1900 nm corresponding to water (Figure 5B, C): the high and positive VIP scores in that region indicate these variables as particularly
important for the model, while the corresponding high and negative regression coefficients indicate a negative correlation between this region and the analyte. Taking into account that this region lies mostly outside the expected important spectral region for phosphates (highlighted in blue in Figure 5), this suggests an inverse relation between the absorbance in that region and the phosphate content.

To require all the models to focus on the spectral regions with the highest correlation to the two analytes, a new set was built using only the most relevant intervals. The intervals were selected based on literature [11,16,45-49] and inspection of the correlation coefficient pictured in Supplementary Figure S1, and were 1600-1900 nm for phosphates and 2050-2300 nm for ammonium. Note that, for phosphates, the selected regions intentionally excluded the peak of water and therefore did not match with the areas of highest VIP scores shown in Figure 3. The performance of these last models based on the prediction error [36], together with those on the whole spectra, are shown in Figure 4 and summarized in Supplementary Table S5.

The models using the intervals showed similar performance to those using the entire spectral region, with prediction errors of similar magnitude, except for those in water (Figure 4). As an example, the models using the intervals for the complex media samples are shown in Supplementary Figure S2.

As highlighted in Figure 5, the model for ammonium focused on the peak reported in the literature [11,16,45,48,49] and highlighted in Figure S1, whereas that for phosphates was based only on the tail of the peak at 1900 nm and with a negative correlation coefficient.

The model for phosphates (and to a lower extent for ammonium) in water had an extremely high error compared to the others (Figure 4). Based on detailed investigation of the prediction error contributions (Appendix B), this was attributed to an error of the model itself, resulting from deviation from the Lambert-Beer law, non-linearity, or generally, a lack of information in the spectral data. This supported the previous hypothesis of lack of information in NIR data (built of the model on indirect
correlations) in particular with respect to phosphate and to a lower degree for ammonium. Apparently, a more complex background than water is needed to gather spectral data rich enough in information to build a PLS model. This is confirmed by the decreased error for NIR models based on the samples of minimal and complex media (Appendix B). Moreover, a comparison in sensitivity of the techniques (Supplementary Figure B1, Supplementary Table S3), highlighted a much lower sensitivity of NIR spectroscopy for both analytes. Comparing the best models (selected among those built on the whole spectra or on the intervals) on MIR and NIR data (Supplementary Table S6), it was observed that the best were always those based on MIR spectra, which had overall the lowest RMSEP with a lower number of latent variables. This is unsurprising considering the lower sensitivity of the NIR and the hypothesis that most models built on NIR spectra were based on indirect correlations. In this regard, basing a model on indirect correlations does not, per se, hamper the possibility of its use: it was observed in Table S6, for example, that some of the NIR models are similar in performance to those built on MIR and could potentially be used to predict the analytes. In the case of models based on indirect correlations, however, one has to bear in mind that the persistence of such correlation becomes an essential condition for future applicability. Any change of the matrix or in fermentation conditions could potentially break the correlation, rendering the model inaccurate. Hence, a frequent and accurate procedure of calibration maintenance becomes of prime importance.

As noted, in the majority of studies where an on-line quantification of ammonium in fermentation medium was attempted, it was by NIR spectroscopy [4,5,9,11,12,14,16,20,21,24,26,48] due to its better characteristics of implementation in an industrial plant. It is never applied to phosphate quantification due to the lack of specific strong signals for this analyte in this spectral region. The performances reported for ammonium vary highly (between about 2 and 25% prediction error) depending on the specific application and method of implementation [4,5,9,11,12,14,16,20,21,24,26,48]. As with MIR spectroscopy, the present results align with some
of the most successful examples both with respect to model performances in prediction, and with respect to the spectral regions used (the region above 2000 nm).

**MIR-Spectroscopy – PLS Models of all samples**

In order to demonstrate the reliability of the approach described, a series of models was computed where the samples of the three solutions were all considered together. They were computed for both analytes, on the whole spectra and on the selected regions. Models based on the MIR data were shown to be superior to the NIR based models as summarized in Supplementary Table S7 and Figure 4E and Figure 6.

In Figure 4E, the overall model for ammonium showed a slightly higher error than that for phosphates, attributed to the lower sensitivity of the technique for the ammonia peak (resulting in very low intensity), compared to phosphates (Figure 2B, C) (see also Appendix B, Figure B2 and related text).

Figure 6 shows the model built on all MIR spectra acquired in the three different media built on the whole region. Focusing on the models based on the complex medium samples representing the approach closest to a real fermentation medium, a comparison of the models in Figure 4A, B with 4E shows how similarly both approaches perform in general. The similar performance independent of the data set used (either combining the three data sets or keeping them separately) is further indicated by comparing the errors of prediction in Figures 6 and 3, in both cases about 3-4%. This confirms that the strategy is robust even where different batches of sample are analyzed together, given that all kinds of samples are present in the calibration set.

Inspecting Figure 6B, C in more detail, it can be seen that a broader spectral region was required to build the PLS models for phosphates than the corresponding one in Figure 3, extending the spectral range found by visual inspection and according to [11,16,45-49] (highlighted in blue). In addition, with regard to the VIP index in Figure 3, Figure 6B shows a more complex pattern in the respective
spectral region. This confirmed the complex nature of the spectral region associated with phosphates and its strong dependence on the medium. In contrast, the model for ammonium presented a pattern of VIP and regression coefficients (Figure 6E, F) that resemble those in Figure 3E, F suggesting a strong correlation in the particular region with the analyte.

**CONCLUSION**

In this paper different spectroscopic techniques have been evaluated with respect to their ability to determine quickly the content of phosphates and ammonium in a wide range of concentration in fermentation media with a minimum of sample preparation. Due to their potential for on-line applications, the performance of NIR and MIR spectroscopy combined with PLS modeling was explored. As reference methods a traditional colorimetric assay kit was chosen for ammonium and $^{31}$P NMR measurement for phosphates. The PLS models based on MIR gave good results, with prediction errors <5% for both analytes in the datasets of the complex and minimal media samples, and even better results for that in water (error ~2%). When PLS models were built on the spectra of all types of samples, the errors remained stable between 3.5 and 5% for both analytes, confirming the generalizability of this approach. The PLS models based on NIRS spectroscopy, on the other hand, provided acceptable results for the concentration of ammonium in some media, and satisfactory but indirect models for quantification of phosphates in the same medium. This method is therefore considered inferior to the previous one for the task under examination even though, given the much easier applicability of the technique on-line, it could be still used under controlled circumstances and following an accurate routine of calibration maintenance.

The present work is one of the few studies attempting and succeeding in simultaneous quantification of two important analytes, phosphates and ammonium, and the only one dealing with such a broad range of concentrations of both. The results are promising in terms of fast and accurate analysis of
phosphates and ammonium in fermentation samples, especially considering that the models are built on experimentally designed samples where the two compounds were uncorrelated and care was put into keeping the models free from dependence on any indirect or spurious correlation.

**FUNDING:**

This work was supported by the BIOPRO project (www.biopro.nu) thanks to the funds granted by Innovationsfondens [Grant Number 4105-00020B]. The funding source had no involvement in the study design, collection, analysis, interpretation of data, writing of the manuscript, nor in the decision to submit the article for publication.

**REFERENCES**


GLOSSARY OF CHEMOMETRIC AND SPECTROSCOPIC TERMS

Chemometrics: scientific discipline of extracting information from chemical data by means of mathematical and statistical models.

Multivariate data analysis: Statistical and mathematical approach to data analysis that implies the simultaneous use of several measured variables on the same set of samples.

Partial Least Square (PLS): regression models: one of the most general and common regression models in chemometrics. For a detailed explanation of its mathematical description see the reference cited in the text [33-35].

Cross Validation (CV): iterative process used as internal validation and to select the most appropriate parameters for the chemometric model. Iteratively, predefined data set segments are left out from the modelling phase and are used as “external samples” to evaluate the performance on the model. These predictions are used against the reference values to estimate the RMSECV.

Root mean Square Error (RMSE): corresponds to the random uncertainty of the model.

Coefficient of determination ($R^2$): proportion of the variance in the dependent variable ($y$) that is predictable from the independent variables ($X$).

Variable Importance in Prediction (VIP): index associated to each variable and explaining its importance in the chemometric model. For a detailed explanation of its mathematical description see the reference cited in the text [44].

Bias: corresponds to the systematic error in the model.

Regression Coefficients: vector of proportionality constants used to describe the relations between the dependent variable ($y$) and the independent variables ($X$). For a detailed explanation of its mathematical description see the reference cited in the text [33-35].
Leverage: diagnostic measure describing the influence of each sample (or each variable) used for calibration, on the mathematical model.

Full factorial design: (or fully crossed design) type of experimental design that includes experiments with all possible combinations of all the factors considered in the design at all their possible levels.

Attenuated total reflectance (ATR): sampling spectroscopic technique that enables the analysis of solid or liquid samples without further preparation or isolation of their components. For a detailed explanation see specialized literature [50].

Mean centering (preprocessing): Variable pre-processing consisting of subtracting by each variable their mean over all the samples. For a detailed explanation of its mathematical description see the reference cited in the text [51].

Standard Normal variate (SNV): Variable pre-processing for scatter correction consisting of for each sample, subtracting each variable of the mean of all the variables for that sample and dividing by the standard deviations of all the variables for the same sample. For a detailed explanation of its mathematical description see the reference cited in the text [51].

Multiplicative Scatter Correction (MSC): Variable pre-processing for scatter correction consisting of rescaling and shifting of the spectra according to scalar parameters found by plotting each spectrum towards a reference spectrum. For a detailed explanation of its mathematical description see the reference cited in the text [51].
**FIGURE LEGENDS**

**Figure 1.** Design of Experiment representation. Samples are represented by their number in a position relative to their concentration in phosphates and ammonium (mM). Samples represented in a circle were selected to be part of validation set.

![Design of Experiment representation](image)

**Figure 2.** Raw infrared spectra of (A) water, (D) minimal medium and (G) complex medium solutions. Insets of the most interesting region for phosphate (B, E, H) and ammonium (C, F, I), on the right. Assignment of the main signals for phosphates and ammonium have been included in insets B and C [38, 41].
**Figure 3.** PLS models on complex medium, whole MIR region (800 – 2250 cm\(^{-1}\)). A) Regression model on the whole region for the concentration of phosphates, with (B) corresponding VIP values and (C) regression coefficients. D) Regression model on the whole spectrum for the concentration of ammonium with (E) corresponding VIP values and (F) regression coefficients. The spectral region correlated to the analyte, found by visual inspection and according to literature [11, 16, 45 – 49], is highlighted in blue (panels B, C, E, F): these regions were used in the Interval models.

**Figure 4.** Comparison of all PLS models on MIR (A, B and E) and NIR (C and D) data by their relative Prediction Error in %. Panels A and B: models for phosphates and ammonium on MIR data, both on the whole spectra (800 – 2250 cm\(^{-1}\)) and on the intervals (950 – 1200 cm\(^{-1}\) for phosphates
and 1360 – 1540 cm\(^{-1}\) for ammonium). Panels C and D: models for phosphates and ammonium on NIR data, both on the whole spectra (400 - 2500 nm) and on the intervals (1600-1900 nm for phosphates and 2050-2300 nm for ammonium). Panel E: models for phosphates and ammonium built on the three sets of samples both on the whole spectra (800 – 2250 cm\(^{-1}\)) and on the intervals (950 – 1200 cm\(^{-1}\) for phosphates and 1360 – 1540 cm\(^{-1}\) for ammonium) of the MIR spectrum.

**Figure 5.** PLS models on complex medium, whole NIR region (400-3500 nm). A) Regression model on the whole region for the concentration of phosphates, with (B) corresponding VIP values and (C) regression coefficients. D) regression model on the whole spectrum for the concentration of ammonium with (E) corresponding VIP values and (F) regression coefficients. The spectral region correlated to the analyte according to literature is highlighted panels B, C, E, F: this region will be used in the following Interval model.
Figure 6. PLS models on three sets of samples, whole MIR region (800 – 2250 cm⁻¹). A) Regression model on the whole region for the concentration of phosphates, with (B) corresponding VIP values and (C) regression coefficients. D) regression model on the whole spectrum for the concentration of ammonium with (E) corresponding VIP values and (F) regression coefficients. The spectral region correlated to the analyte, found by visual inspection and according to literature [11, 16, 45 – 49], is highlighted panels B, C, E, F: this region will be used in the following Interval model.