Analytica Chimica Acta 1284 (2023) 342005

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

The role of biospectroscopy and chemometrics as enabling technologies for upcycling of raw materials from the food industry

Ingrid Måge [a,*], Sileshi Gizachew Wubshet [a], Jens Petter Wold [a], Lars Erik Solberg [a], Ulrike Böcker [a], Katinka Dankel [a], Tiril Aurora Lintvedt [a,b], Bijay Kafle [a,b], Marco Cattaldo [a,c], Josipa Matić [a], Liudmila Sorokina [a,d], Nils Kristian Afseth [a]

[a] Nofima - Norwegian Institute for Food, Fisheries and Aquaculture Research, Muninbakken 9-13, Breivika, 9291, Tromsø, Norway
[b] Norwegian University of Life Sciences, Faculty of Science and Technology, 1432, Ås, Norway
[c] Universidad Politécnica de Valencia, Department of Applied Statistics, Operations Research and Quality, 46022, Valencia, Spain
[d] University of Oslo, Department of Chemistry, 0371, Oslo, Norway

HIGHLIGHTS

- Spectroscopy and chemometrics are crucial tools for understanding bioprocesses.
- Raw material properties and process settings both affect the end-product quality.
- NIR and Raman spectroscopy may be used for in-line characterisation of raw materials.
- FTIR, SEC and NMR provide detailed characterisation of protein hydrolysates.
- Transition of technology and insights from lab to industry is a remaining barrier.

ARTICLE INFO

Handling Editor: Prof. L. Buydens

Keywords:
Enzymatic protein hydrolysis
Raw material variation
Bioprocessing
Spectroscopy
Process analytical technology
Process modelling

ABSTRACT

It is important to utilize the entire animal in meat and fish production to ensure sustainability. Rest raw materials, such as bones, heads, trimmings, and skin, contain essential nutrients that can be transformed into high-value products. Enzymatic protein hydrolysis (EPH) is a bioprocess that can upcycle these materials to create valuable proteins and fats. This paper focuses on the role of spectroscopy and chemometrics in characterizing the quality of the resulting protein product and understanding how raw material quality and processing affect it. The article presents recent developments in chemical characterisation and process modelling, with a focus on rest raw materials from poultry and salmon production. Even if some of the technology is relatively mature and implemented in many laboratories and industries, there are still open challenges and research questions. The main challenges are related to the transition of technology and insights from laboratory to industrial scale, and the link between peptide composition and critical product quality attributes.

1. Introduction

Utilization of the whole animal is critical for the sustainability of meat and fish production. Large volumes of rest raw materials such as heads, bones, trimmings, and skin are created when producing meat and fish products. These fractions contain proteins and other essential nutrients and may be transformed into high-value products for human consumption, pet food, or feed. During the last twenty years, the food industry has invested heavily in bioprocesses that can upcycle these raw materials to nutritionally valuable proteins and fats. One such bioprocess is enzymatic protein hydrolysis (EPH), which is widely used in the production of e.g. sports nutrition and infant formulations from dairy raw materials [1,2] and has recently been adapted to refining peptides from meat and fish rest raw materials [3–7].

In EPH, proteolytic enzymes break down large and undigestible proteins into smaller proteins, peptides, and amino acids. Some raw material sources contain highly active endogenous enzymes, but commercial enzymes are usually added to obtain higher yield and specific

* Corresponding author. Nofima AS, Osloveien 1, N-1430, Ås, Norway.
E-mail address: ingrid.mage@nofima.no (I. Måge).

https://doi.org/10.1016/j.aca.2023.342005
Received 29 March 2023; Received in revised form 25 September 2023; Accepted 5 November 2023
Available online 8 November 2023

product properties. The process starts by grinding the raw materials before they are mixed with water and enzymes. The hydrolysis reaction can take place in reactors in batch, semi-continuous or continuous mode. After 30–60 min of hydrolysis time at an appropriate temperature for the enzyme in use, the reaction mixture is thermally inactivated, and three crude fractions are recovered: a protein-rich liquid phase, an oil phase, and a mineral-rich sediment. In this paper we focus on the protein-rich liquid phase, also called the hydrolysate, which may be further refined by e.g., separation into fractions with specific molecule size intervals. The final product is usually in the form of a liquid concentrate or a powder that can be used as an ingredient, a functional food, or a nutraceutical. A simplified sketch of the process is shown in Fig. 1.

Uncontrollable variation in raw material quality is a general challenge for the food industry, due to differences between seasons, breeds, origins, preprocessing conditions, etc. This is also the case for the EPH industry, as it must process whatever is left after taking out the more valuable parts of the animal. The proportions of fat, bone, connective tissue, and residual meat depend heavily on the cutting/fileting procedures, which may vary within a day, a week, and over the year. In addition, raw materials from poultry processing are associated with substantial differences in composition depending on the type of fowl (i.e., chicken or turkey) [8].

Key performance indicators of the process are the *protein yield (*also called *protein recovery)* and the *degree of hydrolysis.* The yield is the proportion of protein in raw material that is retained in the hydrolysate, while the degree of hydrolysis is the percentage of peptide bonds that have been cleaved in the reaction. The product quality is to a large extent given by the *peptide composition* i.e. the relative amounts of different proteins, peptides, and amino acids in the product. The yield, degree of hydrolysis, and peptide composition are however not independent, as higher reaction time increases the yield and the degree of hydrolysis and leads to a higher content of small molecular weight peptides. Additionally, enzyme type and dose highly affect all three characteristics. Many product properties such as taste, functional properties, digestibility, allergenicity, and bioactivity are directly linked to specific molecular weights and sequences of peptides [9–11]. It is therefore important to be able to control and tailor the peptide composition depending on the desired attributes of the final product, and at the same time optimise yield and profitability.

The product quality is bound to be a function of raw material properties and processing (Equation (1)). Variation in raw materials will therefore lead to variation in the product quality unless the process is adjusted to mitigate the raw material variation. To reach the goal of stable and tailored product quality, it is necessary to understand or model the relationships between raw material properties, process parameters and end product quality. A good model (or approximation) of these relationships can be used to optimise the process and achieve the desired product quality.

$$Product = function(Raw\ materials, Processing) \quad (1)$$

There are several challenges to reaching this goal. The first challenge is to collect enough relevant data to fit such a model. For raw materials, it is not fully known which properties are important, and a comprehensive characterisation is therefore necessary. Regarding the process, it is necessary to collect data from a range of different conditions, by varying for instance enzyme types and doses, temperatures, pH, and reaction times. As for the product, a typical hydrolysate is a complex mixture of a multitude of different peptides, and advanced spectroscopic and chromatographic methods are needed to characterize it. The data that go into the model are therefore highly multivariate, on both sides of the equals sign, creating a need for multivariate data analysis methods and particularly dimension reduction techniques. Additionally, the modelling strategy may be purely data-driven or physics-informed (i.e. hybrid), and different methods are needed depending on the research question or industrial optimisation, monitoring, or control application.

This paper concentrates on the role of spectroscopy and chemometrics in valorising rest raw materials through enzymatic protein hydrolysis. The novelty lies in connecting bits and pieces of previously published research into a bigger picture, with emphasis on the challenging leap from basic research to industrial innovations. To do this we present recent developments in chemical characterisation and process modelling, accompanied by new data and results from industry-scale processing. Our focus is on the rest raw materials from poultry and salmon production, but the analytical techniques are equally relevant for other raw materials of animal, marine, or plant origin. The paper is also relevant for other types of (bio-)processes, as it showcases how spectroscopy and chemometrics can lead to a better understanding of any complex process.

The paper is organized as follows: In sections two and three, we present analytical methods for characterising raw materials and product quality, respectively. In section four, we review the most relevant data analytical methods for modelling relationships between raw materials, processing, and product quality. Finally, we discuss the state of the art and point out directions for the future in section five.

2. Raw material quality

The raw materials typically consist of carcasses with varying amounts of residual muscle, heads, and skin. The relative proportions of these fractions vary, leading to a large variation in gross chemical composition. The overall protein content in the raw material is important for calculating the yield, and for optimizing the enzyme dosage. Also, the amount of collagen is important as collagen affects the functional properties of the final hydrolysates and might also affect the actual processing. The fat content is important since it directly affects the protein yield and since high fat-levels have been shown to reduce enzyme activity in specific processes [12]. The ash content mainly reflects the amount of bones, which contain soluble minerals that should usually be kept below specification limits which may vary between different product categories. If the mineral content is high, the process may require additional downstream processing (i.e., desalting) to achieve a high protein content and satisfactory product quality. Also, the combination of fat, protein, collagen, and ash may indirectly reflect the raw material type (such as skin, carcass or trimmings) and thereby correlate with other properties that are hard to measure directly. Since the main components sum to 100%, the protein and fat content is usually

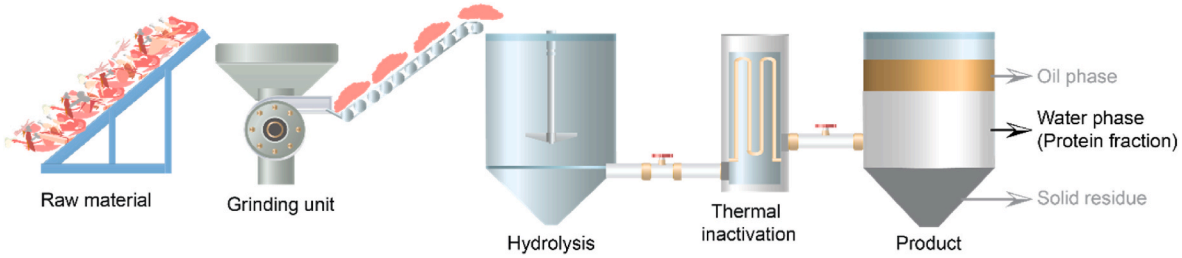

Fig. 1. Sketch of industrial enzymatic protein hydrolysis of rest raw materials from the food industry.

strongly negatively correlated, and it is therefore difficult to independently assess their effects on the process and product quality. The huge variation, not only in gross chemical composition but also in the raw material properties ranging from bone via skin, cartilage, meat and heads, poses challenges for spectroscopic methods that in general are sensitive to variations in texture, particle size, coarseness, pigments etc.

The concept of feed-forward control in industrial EPH of poultry raw materials was introduced by Wubshet et al. [13]. They showed that NIR, fluorescence and Raman spectroscopy had the potential to quantify the contents of ash, fat, and protein in raw material samples and that the information to some extent could be used to predict the quality properties of the end products. Below we describe the potential and status of these three spectroscopic methods regarding efficient in-line monitoring of critical raw material properties.

2.1. Near-infrared spectroscopy

Near-infrared (NIR) spectroscopy is a potent method for industrial quality control and process monitoring, and the potential for process optimisation is large [14]. NIR spectroscopy is an excellent tool for rapid and nondestructive determination of fat, water, and protein in ground meat, by measuring molecular vibrations involving hydrogen bonds (e.g. C–H, O–H, and N–H). This application is well established for both at-line and in-line monitoring in the meat and fish processing industry, and different types of commercial near-infrared instruments are used for this.

Representative sampling is often a challenge when measuring heterogeneous foods with spectroscopic methods. A five-filter NIR reflection instrument was introduced as an in-line method for determining fat levels of batches of ground meat [15]. It works well since an average value can be recorded over a large volume at the grinder outlet. NIR technology based on high-speed hyperspectral imaging in the spectral range 760–1040 nm in combination with so-called interaction measurements enables sampling across the entire width of a conveyor belt as well as about 10- to 15-mm depth into the material. This produces good estimates of fat content in portions of meat trimmings and fish fillets on a conveyor belt [16,17]. This technology enables the automatic sorting of trimmings according to chemical composition [18,19] and could

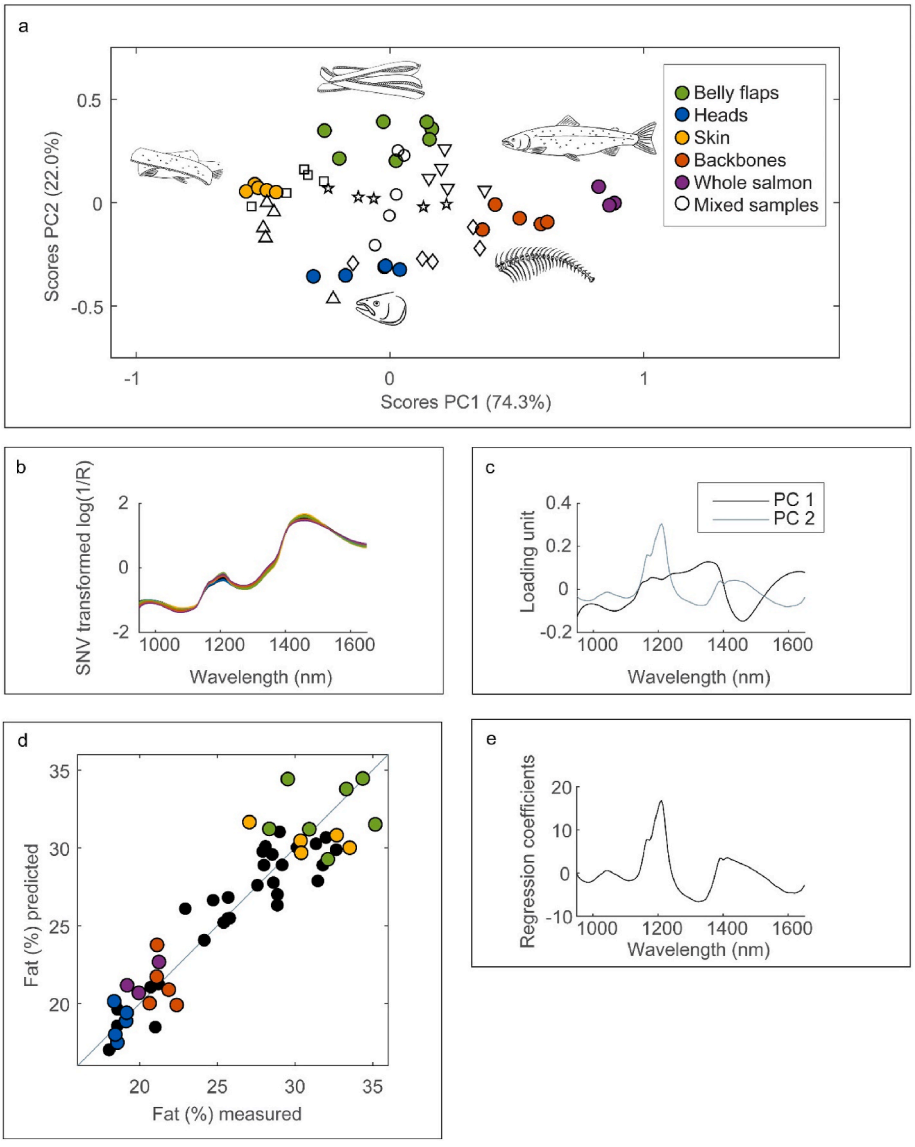

Fig. 2. Results from characterisation of different ground salmon raw material categories by near infrared spectroscopy. (a) Score plot from a PCA of SNV transformed NIR spectra of whole salmon, backbones, skin, heads, belly flaps and mixtures of the listed categories. (b) The NIR spectra presented as standard normal variate (SNV) transformed apparent absorption (log (1/Reflectance). (c) PCA loadings corresponding to the score plot in panel a. (d) Cross-validated predicted versus measured fat content from a calibration model. (e) The regression vector from the PLS calibration model for fat content.

potentially also improve the utilization of the rest raw materials.

It has been shown that in-line NIR "point" measurements in the wavelength range 950–1650 nm has the potential to determine fat, protein, and collagen with acceptable accuracy in heterogeneous raw material samples from both poultry and fish [13,20]. Fig. 2 illustrates the huge variability in the composition of the rest raw materials from Atlantic Salmon (*Salmo salar*). A score plot from a principal component analysis (PCA) of the NIR spectra (reflection in the range 950–1650 nm) from 55 coarsely ground samples shows that the main categories of raw materials spanned the two first principal components, explaining more than 96 % of the variation. PC1 most likely expressed variation in light scattering and varied from skin to ground whole salmon, while PC2 was related to fat content, spanning from lean heads to the very fatty belly flaps. Despite this large raw material variation, it was possible to establish a two-component PLSR calibration model for fat with an acceptable prediction error (RMSECV = 1.7 %). It is important to bear in mind that when such a rough calibration is used in-line, the prediction error is significantly reduced with increasing measurement volume [16]. Since the main raw material categories from salmon have very different properties, an alternative could be to classify the type of raw material entering the process and use this information together with predictions of fat, protein and/or ash.

2.2. Raman spectroscopy

2.2.1. Raman spectroscopy

Raman spectroscopy is becoming a viable tool within process analytics due to recent technological advances. The potential of in-line Raman applications has been demonstrated in several areas, such as in the pharmaceutical and bioprocessing domains [21]. The method can capture subtle chemical distinctions in foods such as fatty acids in muscle foods [22,23], protein structure related to water-holding capacity in pork meat [24] and bone contents and collagen in raw material slurries [13,20].

Obvious challenges for in-line Raman measurements of food materials are the rather low signals from food components as well as a very small sampling area/volume, which is a clear limitation when sampling heterogeneous foods. However, Andersen et al. [25] demonstrated that a wide area 785 nm laser excitation (D = 6 mm) in conjunction with surface scanning can overcome the obstacle of heterogeneity and produce similar results as NIR for protein and fat in heterogeneous raw material samples from salmon and poultry processing. The wide area illumination probes also make measurements less sensitive to variation in working distance than traditional confocal instrumentation [26,27]. This is an important prerequisite for in-line measurements, where such variations often occur. Furthermore, Lintvedt et al. [28] have shown that residual bone concentration in mechanically recovered ground chicken samples can be quantified by in-line scanning using an exposure time of just 4 s. Raman has recently been tested under industrial conditions for in-line monitoring of protein, fat, collagen, and bone residues in ground poultry rest raw materials with promising results [29]. The two latter studies were conducted with a 785 nm wide-area illumination (D = 3 mm) and a stand-off probe positioned about 10 cm above the raw material stream.

An advantage of Raman compared to NIR spectroscopy is that the spectra are more chemically specific so calibrations on complex food materials are simpler and easier to interpret [30]. Raman calibrations might therefore be more robust towards raw material variations in texture, particle size and temperature [22], making it a good candidate for raw material characterisation despite the above-mentioned limitations. It is also well known that Raman scattering from water is very weak, which means that signals from water do not critically obscure bands from other analytes, which can be a challenge with NIR spectroscopy for instance.

2.3. Fluorescence spectroscopy

Fluorescence spectroscopy is a sensitive method that lends itself to in-line monitoring of foods. The method can be highly specific, but spectra from meat and fish are composed of largely overlapping signals from different chromophores. It has been shown that collagen (hydroxyproline) can be determined in minced beef [31], oxidation in minced poultry [32] and protein in poultry slurry [13] and the method can distinguish between e.g. collagen and elastin [33]. These methods rely on excitation in the 330–380 nm wavelength range and collection of the informative fluorescence emission spectra in the visible region (400–780 nm). A challenge with the technique with respect to rest raw materials is that the fluorescence is reabsorbed by pigments (i.e. the inner filter effect), meaning that varying amounts and forms of e.g. myoglobin can obscure the spectra and hamper quantitative modelling [20].

2.4. Practical aspects of spectroscopic methods in the food industry

Grassi and Alamprese [13] point out that there is a gap between promising laboratory studies with NIR spectroscopy and actual process analytical implementations in the food industry. The complexity of a food process represents a challenge to the implementation of spectroscopic sensors. The sensors can be expensive, and they usually require calibration and maintenance over time by experienced persons. Nevertheless, there are many well-functioning in-line systems in the food industry. In some processes, these solutions are very important for sustainable and profitable operations. It all becomes a question of benefit in relation to cost.

3. Product quality

An EPH process, and not least the outcome of an EPH process, can be characterised in several different ways. One of the most industrially relevant properties is the protein content, which is necessary for calculating protein recovery. Moreover, depending on the intended application, the peptide composition must meet a set of specifications. For example, the average molecular weight (AMW) of protein hydrolysates has been associated with important quality attributes including nutritional profile, therapeutic value, and functionality. The degree of hydrolysis (DH) is also a well-established parameter, both for describing the extent of hydrolysis in the resulting peptide product and for monitoring the EPH process itself. It has been shown that DH and AMW are highly correlated parameters [34]. Thorough overviews of classical analytical approaches for protein content and degree of hydrolysis assessments can be found elsewhere [35,36]. In this section, we focus on non-destructive and spectroscopic approaches.

3.1. Protein content and yield

3.1.1. °Brix

The °Brix value represents the amount of dissolved solids in a solution. It is used extensively in the food industry, mainly to measure sugar content in fruit and vegetable products. °Brix is usually measured with a refractometer, which is cheap and easy to use and may also be implemented inline [37]. The main components affecting °Brix are carbohydrates and proteins. Protein hydrolysates contain low amounts of carbohydrates, and the °Brix value may therefore be used as an estimate of protein content. We have experienced that there is a strong linear relationship between °Brix and protein content and propose that °Brix is a candidate for a low-cost indirect measure of protein yield either at-line or in-line.

3.1.2. Near-infrared (NIR) spectroscopy

Due to the benefits of NIR in gross component food analysis, the technique is especially suited for inline documentation of fat, protein,

and water contents of dried or concentrated protein hydrolysates and sediments. NIR is also a promising technique for in-situ monitoring of the actual hydrolysis process. Using a miniature NIR spectrometer, Zhang et al. showed that NIR is well-suited for in-situ and real-time monitoring of the enzyme-assisted hydrolysis of wheat gluten [38]. The same research group also showed that by using the sulfhydryl groups and disulphide bond information of the NIR spectrum in wheat gluten hydrolysis, the effect of ultrasound on increasing the hydrolysis efficiency could be studied [39]. However, in batch processes where the protein content of the liquid phase is bound to increase during hydrolysis, it is important to be aware of the close relationship between DH and protein concentration over time, indicating that NIR-based estimation of DH could be done based on indirect modelling of the protein yield [38].

3.2. Protein quality

3.2.1. Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) coupled to a UV detector is the main analytical technique used to assess the molecular weight distribution (MWD) of protein hydrolysates. However, a technique such as SDS-PAGE has also been presented as a complementary analytical tool to provide insight into the MWD of hydrolysates [40]. The raw chromatograms from SEC may be evaluated directly, as detector response versus retention time, or various MWDs may be derived by using known calibration standards [41,42]. Many prefer the *differential log molecular weight distribution,* x(M), which uses the logarithm of molecular weight as a basis (see Fig. 3 for an example). The x(M) is especially useful for samples containing fractions of very different molecular weights, as is often the case for hydrolysates of complex raw materials. Parameters such as weight average molecular weights (AMW) and relative fractions of specific molecular weight ranges can also be derived from the weight distributions.

SEC has been used both for product characterisation and for process monitoring. In addition to AMW, it is often necessary to assess the full MWD to fully understand the hydrolysate composition from a given process. This is exemplified in Fig. 3, which shows the MWD of two industrial hydrolysates that have approximately the same AMW but different distributions. Integrals of five molecular weight ranges were calculated and their numbers are given in Fig. 3. Such differences may be unimportant if the end product is feed, but highly relevant if the end product is a functional food or nutraceutical considering that different molecular weight ranges have been found to correlate with bioactivity [11]. AMW alone is therefore not sufficient to characterize hydrolysate quality.

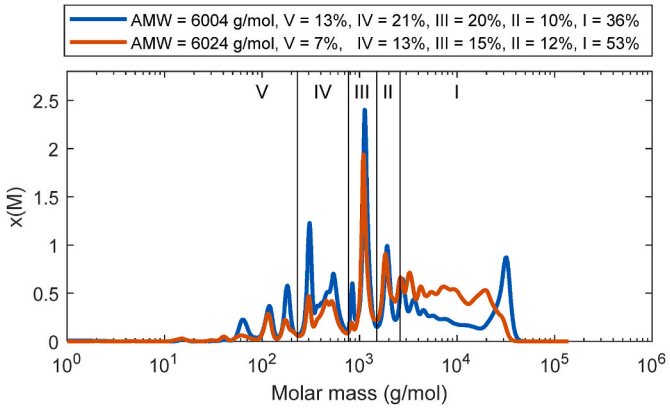

Fig. 3. Differential log molecular weight distribution, x(M), of poultry hydrolysates, derived from size exclusion chromatography. The plot shows two hydrolysates with approximately equal average molecular weight (AMW) but different weight distributions (MWD). The distributions are divided into five molecular weight regions (I-V) as in Ref. [11], and the relative weight proportion of each fragment is calculated from the cumulative weight distribution.

While SEC is a powerful characterisation platform for protein hydrolysates, it has critical limitations related to poor resolution and narrow exclusion range. In addition, samples eluting close to the permeation limit can often be affected by secondary interactions. Ideally, separation by SEC should solely be based on the size of the molecule. However, secondary electrostatic and hydrophobic interactions of proteins and peptides with the stationary phase is a common phenomenon and can lead to uncertain molecular weight estimations [43]. Furthermore, achieving efficient chromatographic separation requires optimisation of the flow rate and temperature, as well as the addition of organic solvents or pH modifiers to the mobile phase. These can affect the native structure of the protein, and thus cause inaccurate MWD estimations, especially for larger peptides and proteins.

3.2.2. Infrared (IR) spectroscopy

FTIR is an established method for the characterisation of protein and polypeptide structures [44]. Thus, the use of FTIR spectroscopy for probing protein modifications and changes due to the action of proteolytic enzymes is also well documented [45]. Poulsen et al. showed that FTIR can be used to predict the degree of hydrolysis (DH%) during whey protein hydrolysis [46]. Recently, an FTIR-based multivariate approach for monitoring the change in AMW during enzymatic hydrolysis of chicken by-products and whey has been reported [47,48]. These studies revealed that amide absorptions (i.e., amide I at ~1650 cm−1), NH3+ deformation (1516 cm−1), and COO- stretching (1400 cm−1) are important for the prediction of AMW or DH%. From these studies, it is also apparent that protein hydrolysates originating from different raw materials and even different enzymes will display different FTIR fingerprints. This raw material effect may be handled using a hierarchical modelling approach [49]. Kafle et al. showed that FTIR can be used to predict protein quality features of industrially produced protein hydrolysates, thus proving that the technique has sufficient sensitivity to predict protein quality variations that are highly relevant for industrial purposes [50].

Even if several studies have shown that FTIR can be used to predict AMW, the relationship between FTIR and specific molecular weight ranges has not been presented before. We fitted PLS regression models for the five weight ranges shown in Fig. 3, using FTIR spectra in the range 1700–800 cm^{-1}. The spectra were acquired and pre-processed as described in Ref. [51]. The models were based on 50 poultry-based hydrolysates produced in a small-scale lab setup. Five different enzymes and two reaction times were used, creating a wide span of molecular weight distributions, with AMWs ranging from 3400 to 17,500 g/mol. For each model, the optimal number of components was selected based on 10-fold random cross-validation. The models were then tested on 139 hydrolysates collected from full-scale industrial production, with an AMW range of 3000–8000 g/mol. The industrial hydrolysates were produced from the same type of poultry raw material, but with a different enzyme and with different processing conditions. Details on model parameters, cross-validation and test set metrics are given in Table 1 and predicted versus reference values from cross-validation and test set are given in Fig. 4. The models predicted the test set surprisingly well, although a bias correction was needed for AMW and ranges II and III. This is probably due to uncertainties in the molecular weight distributions, caused by differences between SEC columns.

The FTIR approach is not limited to the quantification of molecular weight characteristics alone. Sorokina et al. [11] recently showed that the FTIR fingerprints can be correlated to selected bioactivities of poultry protein hydrolysates, and Kristoffersen et al. showed that FTIR can predict collagen content in hydrolysates [52]. We have also demonstrated that the FTIR fingerprint of hydrolysates can be used directly, without calibration towards specific references [51]. The paper showed that the FTIR fingerprint, decomposed by PCA, could be related to differences in enzyme action, processing time and raw material

Table 1
Evaluation of PLS regression models predicting AMW and relative proportions of five molecular weight ranges (all derived from SEC), using FTIR spectra as predictors. RMSE = Root Mean Squared Error, SEP_b = bias-corrected Standard Error of Prediction. Note that $RMSE^2 = bias^2 + SEP_b^2$. Predicted versus reference values from cross-validation and test set is shown in Fig. 4.

	# PLS components	R^2_{CV}	$RMSE_{CV}$	$RMSE_{test}$	bias	$SEP_{b,test}$
AMW (g/mol)	4	0.86	1288	5159	−5078	913
Range V, <230 g/mol (%)	3	0.87	0.9	1.0	−0.5	1.0
Range IV, 230–770 g/mol, (%)	4	0.92	1.2	1.2	−0.4	1.2
Range III, 770–1500 g/mol (%)	5	0.89	1.2	1.7	−1.3	1.0
Range II, 1500–2600 g/mol (%)	4	0.82	1.0	4.1	4.0	0.8
Range I, >2600 g/mol (%)	5	0.94	2.9	2.8	0.3	2.8

composition. We also showed that fingerprints of laboratory-made hydrolysates could be used as a benchmark for industrial products. We have recently collected new samples from the same factory, during a period when they tested different enzymes. Samples were measured by FTIR and projected onto the PCA model in the same way as described in Ref. [47], see Fig. 5. Subplot (a) shows the first two principal components, which account for the largest amount of variation and were previously found to be related to differences between enzymes and reaction time [47]. We see that the new samples are separated with regard to enzymes in these components, as expected. Subplot (b) shows the third and fourth components. In Ref. [51], PC4 was found to be related to the amount of fat in raw material, and we see here that it separates turkey raw material (which has more fat) from chicken. This validates the conclusion in Ref. [51], stating that FTIR captures industry-relevant variations and is a promising tool for industrial use. A deeper interpretation of the industrial hydrolysates compared with the PCA is out of scope here.

3.2.3. Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance (NMR) is a powerful spectroscopic tool that allows detailed characterisation and structural elucidation of organic molecules. NMR can provide detailed qualitative and quantitative information about molecule(s) in the form of chemical shifts, coupling patterns and peak areas. A typical proton (^{1}H) NMR spectrum of a protein hydrolysate is a complex chemical fingerprint with several overlapping peaks from peptides, amino acids, and metabolites (see Fig. 6). Assignment and quantification of individual NMR peaks to respective constituents in such complex fingerprints is a challenging task. However, NMR-based multivariate statistics have been demonstrated as a valuable tool for both monitoring the enzymatic hydrolysis process [53,54] as well as the characterisation of the peptide products [55]. In one of the early examples, Sundekilde et al. demonstrated NMR as a suitable analytical platform for at- and inline monitoring of the major metabolite composition changes during enzymatic hydrolysis of chicken muscle [54]. While such a high-field NMR approach is a valuable tool for understanding and optimizing processes on a laboratory scale, its industrial application is expectedly challenging due to high cost, magnet volume, and special acquisition requirements (e.g. solvents). Hence, easy-to-operate, cheaper, smaller, and lighter permanent magnets are continuously being developed to meet industrial demands in bioprocess monitoring [56]. In recent work, a bench-top NMR system has been demonstrated as a promising tool for online monitoring of enzymatic hydrolysis of marine raw materials [53]. Such a benchtop NMR system can be configured to operate near the production environment and therefore holds great potential for industrial application. In addition to the process monitoring, NMR has also been used to characterize and understand the quality traits of protein hydrolysates [55,57].

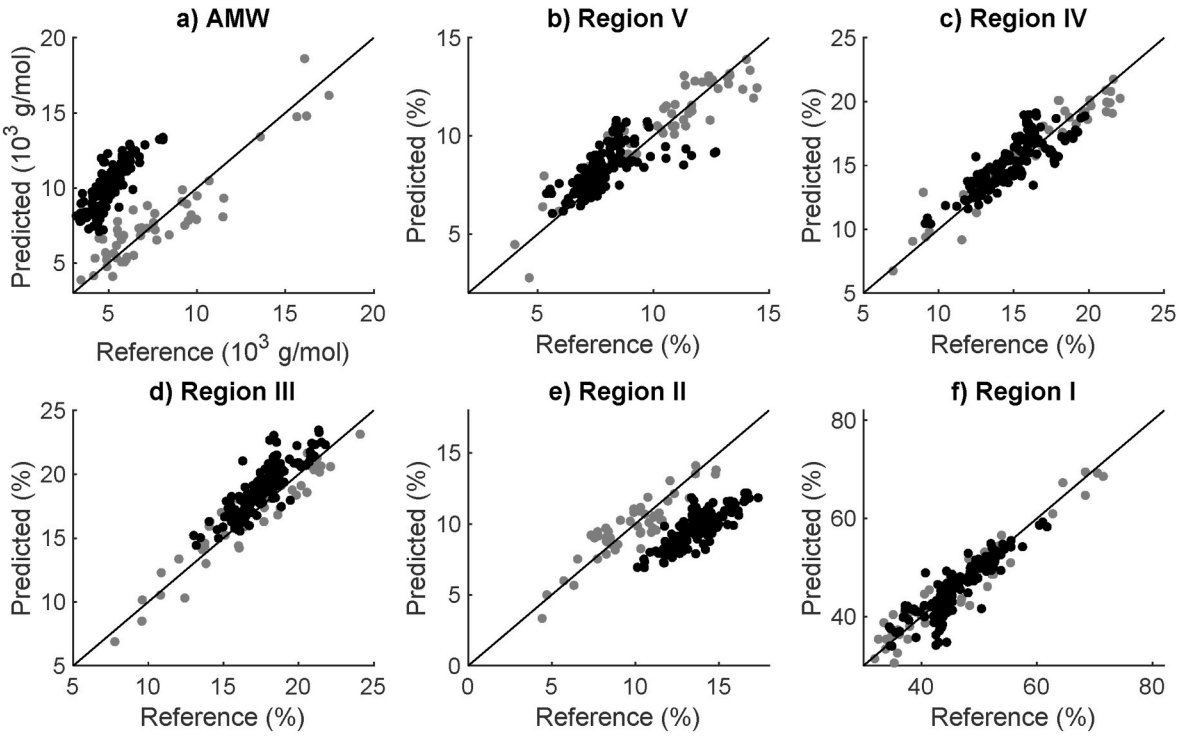

Fig. 4. Predicted versus reference values from PLS regression models based on FTIR spectra. Separate models were made for AMW (b) and the five molecular weight regions (b–f) corresponding to those shown in Fig. 3. The models were optimised by cross-validation using a set of hydrolysates produced in the laboratory (grey dots) and tested on a set of hydrolysates collected from industry (black dots). Detailed model characteristics are given in Table 1.

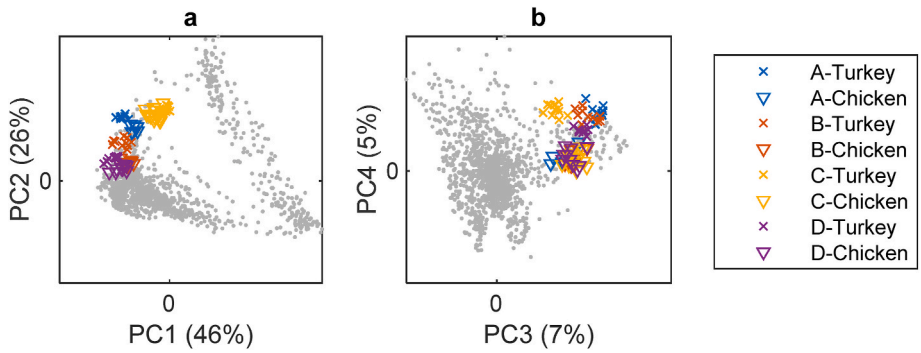

Fig. 5. Hydrolysates from industrial enzyme tests projected on the FTIR PCA model published in Ref. [51]. The grey dots represent the same laboratory-produced hydrolysates that were published in Ref. [51], while the marked samples correspond to newly collected industry hydrolysates based on different enzymes (colours) and raw materials (symbols). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

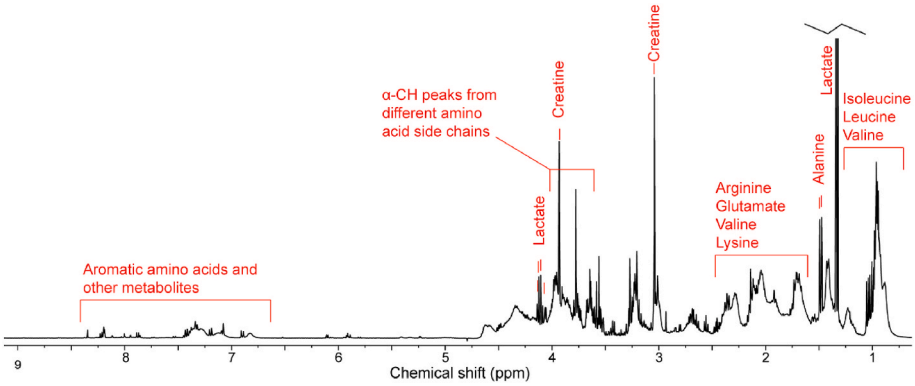

Fig. 6. Representative ^{1}H NMR spectrum of chicken hydrolysate, hydrolysed with FoodPro PNL (0.1 % w/w) for 60 min. Assignments were done for the abundant peaks annotated in red, according to Ref. [54]. Only peaks diagnostic to a given molecule were assigned. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Steinsholm et al. developed ^{1}H NMR-based partial least-squares regression (PLSR) models for prediction of important sensory attributes (e.g., bitterness) for protein hydrolysates derived from cod, salmon, and chicken.

4. Modelling relationships between raw materials, processing, and product quality

Effects of processing factors on protein yield and DH have been studied extensively in controlled laboratory experiments [5,58,59]. The most influential process parameters are enzyme type, enzyme-to-substrate ratio, and reaction time. In addition, temperature control is essential for the enzyme activity. A few experimental studies have also investigated the impact of raw material quality [8,13,47,51]. All these studies show significant effects of both raw material properties and process settings and a model in the form of Equation (1) is therefore needed to understand and predict product quality. There are, however, to our knowledge, no publications that study these effects on an

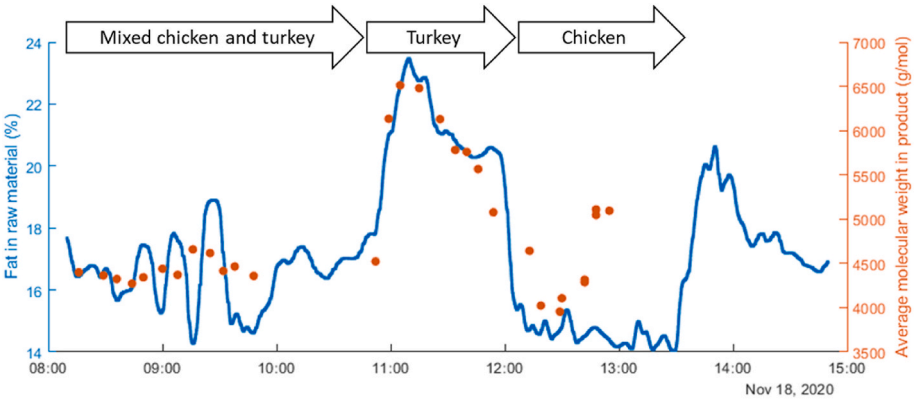

Fig. 7. Industrial measurements show a relationship between raw material composition and the average molecular weight of the hydrolysate. On one production day, the raw material input was changed systematically while the rest of the process parameters were held constant. The blue line shows fat% in raw material, predicted from in-line NIR measurements, while the red dots represent product samples. The average molecular weights of product samples were determined by size exclusion chromatography. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

industrial scale.

Fig. 7 shows an example from industrial hydrolysis of poultry, which is done in a continuous flow process. On one production day, the raw material input was changed systematically while the rest of the process parameters were held constant. The raw material stream was monitored by in-line NIR spectroscopy, while samples of the hydrolysate were taken manually, and AMW was determined by SEC. In Fig. 7, we see that in-line NIR can capture differences between the raw material types and that the change in raw material influences the molecular weight of the hydrolysate. In this particular case, higher fat content also corresponds to higher collagen content, and it is likely that the increase in molecular weight is caused by collagen rather than fat. There are three samples that do not follow the general pattern, around 13:00 in Fig. 7. The reason for this deviation is not known. It could be due to unplanned changes in process settings (e.g. the enzyme-to-substrate ratio), or it could be an error in the registration of sampling time stamps. Unfortunately, we do not have any data to corroborate these hypotheses. This relationship between raw material and hydrolysate quality was not easily recognised on other production days, where variations in raw materials were more random, other process factors were not necessarily held constant, and hydrolysate samples were taken less frequently. More research is therefore needed to fully understand the effects of raw material variation in industrial EPH.

The choice of data modelling method should depend on the analysis goal and properties of the data [60]. The goal may range from extracting new knowledge (through e.g., explorative analysis, identification of important raw material and processing factors, and estimation of causal effects), to real-time process monitoring, control, forecasting and anomaly detection. In general, data can be categorized as *experimental* or *observational* depending on the data collection. Experimental data is collected in controlled settings, often following a statistical experimental design. Observational data is collected from a system under normal conditions and relies on natural variation of the variables. Data produced in the lab are experimental, while data collected from industry can usually be regarded as observational.

Multichannel data such as spectroscopy and chromatography are traditionally modelled by chemometric methods (e.g. PCA, PLS and MCR). The advantages of these methods are that they handle multicollinearity and provide graphical tools, such as scores and loadings/profiles, that can be used to provide chemically meaningful interpretations. Since we characterize both raw materials and products by spectroscopic and chromatographic techniques, we will mainly focus on data modelling methods based on traditional chemometric tools. The next subsections briefly describe the most relevant methods for modelling relationships between raw materials, processing factors and end product quality (Equation (1)) in EPH processes.

4.1. Multivariate ANOVA

ANOVA is the most common family of methods for analysing data from experimental designs. For multivariate and collinear responses, such as spectra, several multivariate ANOVA methods that combine classical ANOVA and latent variable techniques are available. Methods that combine ANOVA and PCA in different ways are most widespread, for instance, ASCA [61–65], ANOVA-PCA [66], ACOMDIM [67], and fifty-fifty MANOVA [68]. There are also methods that replace PCA with PLS regression and thereby utilize PLS-specific validation and variable importance routines [69,70]. The advantage of all these methods in general is that they provide estimates of multivariate effect sizes with corresponding p-values, in addition to well-known interpretation and variable importance metrics from latent variable-based methods.

Multivariate ANOVA has been used to assess the effects of raw material and processing factors on both spectra and chromatograms of protein hydrolysates [47,55]. These papers present multivariate ANOVA tables with information on effect size and statistical significance.

4.2. Multiblock methods

The aim of multiblock modelling is to improve the interpretability of multivariate data when variables can be divided into meaningful blocks, for instance representing different spectroscopic instruments or specific processing steps. The first multiblock methods were presented in the 80's, both for explorative and predictive modelling, and many methods suited for different data types and analysis goals have been developed since then. A recent and comprehensive overview is given in Ref. [71]. One of the most used methods is SO-PLS regression [72] and its many extensions [73–76]. SO-PLS is specifically designed to assess the contribution of different blocks sequentially, i.e. to what extent a block holds new information that improves the model's predictive ability. This property is especially appealing when there is a natural order between blocks, for instance according to time (e.g. different steps in a production line) or ease/cost of data acquisition (e.g. in-line, at-line, and off-line measurements).

SO-PLS has been successfully used to increase understanding of EPH processes. In Ref. [13], it was used to quantify how different raw material characterisation methods contributed to the prediction of protein yield and average molecular weight, while in Ref. [51] it was used to quantify the individual contributions of raw material composition, process parameters, and their interactions on protein quality.

4.3. Soft sensors

A soft sensor, also called a virtual sensor, typically predicts the value of a hard-to-measure key quality variable by using a set of easily available hardware sensors as predictors. In principle, any prediction method may be used to develop a soft sensor. Multiblock regression methods, as described above, are prediction methods but focus on interpretation. It is therefore more efficient to use other methods aimed at prediction, for instance, PLS regression or more flexible machine learning methods. Important tasks when developing a soft sensor include data collection, pre-processing, feature extraction, variable selection, model optimisation, validation, and maintenance. Recent reviews may be found in Refs. [77,78]. Yin et al. [78] point out that most publications are based on process simulators (such as the Tennessee Eastman process [79]) or limited amounts of real industrial data, and that a major challenge is to bridge the gap between simulations and industrial practice. This is also the status of soft sensors in the EPH industry.

A soft sensor for hydrolysate quality would open new possibilities for process optimisation, monitoring, and control of EPH processes. Currently, the protein quality can only be characterised by off-line laboratory methods (see section 3.2), and samples need to be collected manually. Since both raw materials and processing affect the end product properties, a soft sensor would need to use data from in-line sensors on both raw materials and process parameters as predictors. In-line raw material measurements are usually spectroscopic, while in-line process variables are typically temperatures, speeds, torques etc. We have attempted to develop soft sensors for average molecular weight in continuous EPH of chicken and salmon materials, but so far, the amount and quality of data have been the limiting factors.

Preliminary results on a soft sensor for predicting the average molecular weight of salmon hydrolysates are shown in Fig. 8. This example is derived from an industrial continuous flow system, where the entire journey from raw material inlet to the end product spans approximately 4 h. During six non-consecutive production days, data from inline sensors (NIR on salmon raw material, and various temperatures and flow rates) were collected continuously, while the end product (the protein hydrolysate) was manually sampled and analysed by FTIR and SEC. The continuous variables were aggregated within time windows of ten, twenty, or 30 min. The selection of the time window depended on the proximity of the sensor to the end product sampling location. Sensors located farther from the end product required larger time windows due

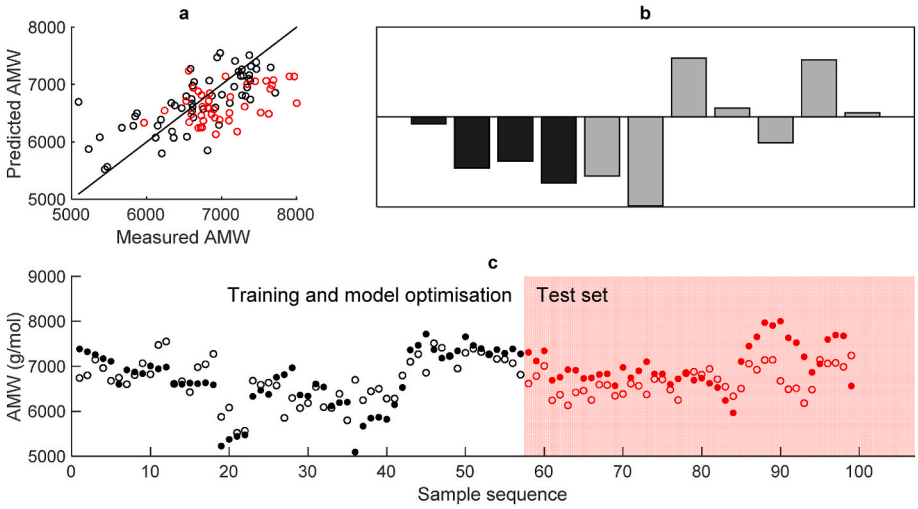

Fig. 8. Preliminary results on a soft sensor for predicting average molecular weight (AMW) of salmon hydrolysates. (a) Predicted versus measured AMW, black circles represent predictions from training (cross-validation) while red circles are test set predictions. (b) Regression coefficients for raw material (black) and process (grey) variables. (c) Timeline of reference values (filled dots) and predictions (open circles) for training and test set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

to some degree of equalization in the process, caused by holding tanks and back-flow mechanisms. Additionally, we employed time-shift adjustments to account for the physical distance between each sensor and the end product. The final set of predictors were four variables related to raw material quality (measured by NIR) and seven process variables. The soft sensor is an ordinary PLS regression model based on 57 training samples collected over the first four production days. The model was tested on 42 samples collected over the two last days. The residual prediction deviations (RPD, defined as the standard deviation of reference measurements in the training set divided by the root mean squared error), are 1.7/1.5/1.2 for training, cross-validation and test set respectively. Predictions from cross-validation and the test set are shown in Fig. 8a) and c), and regression coefficients in Fig. 8b).

Even if the model performance is far from perfect, the test set predictions capture some general time trends (see Fig. 8c), indicating that a well-functioning soft sensor of this type is realistic if more and better data is collected. Note that in addition to the low number of samples, there were challenges with data quality. The NIR measurements were affected by uncontrollable temperature variations in the raw material, as well as back-splash causing a dirty lens. Also, there was little variation and thereby low information content in some process variables due to constant set-points. All the predictors were time series, sampled at different time resolutions and in different locations in the process. In this preliminary work, the aggregation and synchronisation of these were done rather ad hoc, based on information from the process operators. Research has shown that choices of aggregation method (such as interpolation or window-filtering) and metrics (like mean, median, and variation) may have a significant influence on a soft sensor's accuracy [80], and more attention should be given to optimizing this step of the data analysis pipeline in future work. We are currently collecting more and better industry data to continue the development of a soft sensor for hydrolysate quality.

4.4. Statistical process monitoring

Industrial EPH is either run as a batch or continuous process. In both cases, inline measurements provide (multivariate) time series that may be used to monitor the process and detect faults and deviations at an early stage. Statistical process monitoring was introduced 100 years ago by Walter A. Shewhart and has since then been an active field of research and development. Some recent reviews of the field may be found in Refs. [81,82].

End-point detection is a crucial part of monitoring batch processes. In EPH, it is important to stop the process at the correct time to meet product specifications on protein quality. This task is especially challenging since raw materials vary substantially from batch to batch, affecting the optimal reaction time. Several methods for end-point detection in cases with large batch-to-batch variation exist [83,84], but they have as far as we know not been tested in the EPH industry. Some industrial actors determine the end-point by in-line NIR prediction of dissolved protein, but this is not published. Recently, a study that evaluates deep neural networks for forecasting the trajectory of a batch EPH process was published [85]. It uses encoder-decoder networks to predict future FTIR spectra based on spectra taken early in the process. The results are promising, showing that this too is a viable method in cases with large batch-to-batch variation.

5. Discussion and future prospects

The examples used in this paper are limited to the hydrolysis part of the production process. Upstream processing steps such as conditioning and pre-treatment of raw materials and downstream steps such as phase separation and drying also affect the product quality and need to be optimised. For instance, it is known that suboptimal phase separation gives a higher fat content of the hydrolysate, and high drying temperatures trigger the Maillard reaction and may cause a burned taste. Also, laboratory experiments have shown a filtration step before the thermal inactivation may be used to control the collagen content of the hydrolysate [59].

Most of the work in this field is done on a laboratory scale, but more industry-scale trials are coming. Data (amount and quality) is a limiting factor both from laboratory and industrial scale. Laboratory data usually have good quality regarding variation and precision but a low number of samples. Industry data from in-line sensors are abundant but often have little variation and thereby low information content. Also, key quality indicators are often measured at/off-line (or not at all) in the industry, limiting the amount of data points considerably. One of the most promising results for bridging the gap between laboratory and industry is that FTIR characterisation of hydrolysate quality is robust over time and comparable between laboratory and industry samples, as described in section 3.2.2. This means that FTIR is an enabler for improved understanding of product quality in industry, where basic experiments and analytical tools such as SEC and NMR are not available.

So far, we have focused on measuring chemical properties of the

hydrolysate such as protein and fat content, and peptide composition. However, it would be more relevant to quantify the critical quality attributes of the end product directly, for instance, bitterness, solubility, blood sugar regulating capacity etc. Most of the perceived quality attributes are connected to chemical composition, and it is therefore expected that it would be possible to predict them directly from e.g., FTIR and/or NMR spectra. As of today, products for the human market are in the development phase and there are no established targets for quality attributes. More research and development are needed to define relevant quality attributes and their specifications for different end products and to quantify these from analytical measurements.

Results show that hydrolysate quality is not only affected by raw materials and processing individually but also through interactions between the two. This means that the optimal processing conditions depend on the raw material and may need to be adjusted based on specific raw material characteristics. There are three main strategies to tackle this challenge: The first strategy is the *Taguchi* method, dating back to the 1050s. The idea behind this method is to use a design of experiments to optimise both the mean and the variation of the output, ensuring that the selected optimum is robust towards uncontrollable/noise factors (such as raw material quality). The Taguchi method has already found its use in bioprocessing [86] and is a relevant strategy for optimisation of EPH processes. In the second strategy, raw materials are sorted into different quality categories before processing. Then, different process settings may be used for each category, or the categories may be used to produce different products (e.g., pet food, protein supplements or functional foods). This is also known as *prediction sorting* [87,88]. The main statistical challenge with this approach is that it requires precise prediction models, and the main industrial challenge is that sorting and storage facilities for raw materials are needed. The third strategy is real-time process control. If the critical raw material properties are measured continuously at the inlet of the process, parameters such as enzyme load or temperature may also be adjusted continuously. This may be done as part of an automatic control system, or manually by providing decision support to the process operators. Such an approach also requires good estimates of the effect of manipulated variables, in addition to knowledge about dynamics, time delay and dead time across different process steps. Also, this approach may not be feasible if the raw material varies with high frequency, due to mixing and equalization later in the process. In summary, it is not clear which strategy that is industrially most relevant. It depends on statistical, practical, and economic considerations. More work is required to develop control strategies that can be implemented in the industry.

Digital Twin is an emerging technology in many types of manufacturing, including the food industry. In short, a digital twin is a digital representation of a physical object or system. It consists of a network of models, often a mix of geometric, physics-based, data-driven and hybrid models. There are many recent review papers on the topic, see for instance Refs. [89,90]. Typical use cases for digital twins are in the process and product development phase (through simulation and what-if-analyses), and in daily operation (through process monitoring, control, and maintenance). However, the development of digital twins is a challenging task and there are relatively few published cases so far. There are also some specific challenges for bioprocesses: The physiochemical reactions are complex and there is limited knowledge about the underlying mechanisms. Pure mechanistic models are therefore not feasible, and one must turn to hybrid or data-driven models which are non-causal and thereby have problems with extrapolation and robustness [91]. Also, a digital twin needs to be fed with relevant data from the physical world, and for EPH processes this includes data on raw materials and product quality. As described in this paper, such data is not always easy to obtain in industrial settings. Even if there is a long way to complete digital twins for EPH processes, it is realistic to envision twins for specific unit operations or parts of the process. We expect more research on digital twins for EPH and other bioprocesses in the near future.

6. Conclusion

Enzymatic protein hydrolysis of heterogeneous raw materials is a complex bioprocess, and the resulting protein product is a mixture of many different (poly)peptides and amino acids. Advanced analytical methods and chemometrics are crucial for characterising the product quality, and for understanding how raw material quality and processing affect the product. This paper presents recent developments in chemical characterisation and process modelling, showcasing how spectroscopy and chemometrics can lead to a better understanding of complex bioprocesses. Even if some of the technology is relatively mature and implemented in many laboratories and industries, there are still open challenges and research questions. The remaining questions are mostly related to the transition of technology and insights from laboratory to industrial scale and to the link between peptide composition and critical end product quality attributes.

Funding

This work was funded by The Research Council of Norway through the projects SFI Digital Food Quality (project number 309259), SmartBio (project number 282466) and TailoTides (project number 320086), and by the Norwegian Agricultural Food Research Foundation through the project Precision Food Production (project number 314111).

CRediT authorship contribution statement

Ingrid Måge: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization, Supervision, Funding acquisition. **Sileshi Gizachew Wubshet:** Conceptualization, Methodology, Writing – original draft, Supervision, Funding acquisition. **Jens Petter Wold:** Conceptualization, Methodology, Writing – original draft, Supervision, Funding acquisition. **Lars Erik Solberg:** Methodology, Formal analysis, Writing – review & editing. **Ulrike Böcker:** Methodology, Investigation, Data curation, Writing – review & editing. **Katinka Dankel:** Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft. **Tiril Aurora Lintvedt:** Methodology, Investigation, Writing – review & editing. **Bijay Kafle:** Methodology, Investigation, Writing – review & editing. **Marco Cattaldo:** Methodology, Investigation, Data curation, Writing – review & editing. **Josipa Matić:** Methodology, Investigation, Visualization, Writing – review & editing. **Liudmila Sorokina:** Methodology, Investigation, Writing – review & editing. **Nils Kristian Afseth:** Conceptualization, Methodology, Writing – original draft, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We thank Bioco, Norilia, and Biomega for providing raw material and product samples, production data, and expertise on industrial enzymatic protein hydrolysis.

References

[1] A. Nasirpour, J. Scher, S. Desobry, Baby foods: formulations and interactions (A review), Crit. Rev. Food Sci. Nutr. 46 (2006) 665–681, https://doi.org/10.1080/10408390500511896.

[2] J.E. Tang, D.R. Moore, G.W. Kujbida, M.A. Tarnopolsky, S.M. Phillips, Ingestion of whey hydrolysate, casein, or soy protein isolate: effects on mixed muscle protein synthesis at rest and following resistance exercise in young men, J. Appl. Physiol. 107 (2009) 987–992, https://doi.org/10.1152/japplphysiol.00076.2009.

[3] T. Aspevik, Å. Oterhals, S.B. Rønning, T. Altintzoglou, S.G. Wubshet, A. Gildberg, N.K. Afseth, R.D. Whitaker, D. Lindberg, Valorization of proteins from Co- and by-products from the fish and meat industry, Top. Curr. Chem. 375 (2017) 1–28, https://doi.org/10.1007/s41061-017-0143-6.

[4] D. Lapeña, K.S. Vuoristo, G. Kosa, S.J. Horn, V.G.H. Eijsink, Comparative assessment of enzymatic hydrolysis for valorization of different protein-rich industrial byproducts, J. Agric. Food Chem. 66 (2018) 9738–9749, https://doi.org/10.1021/ACS.JAFC.8B02444/ASSET/IMAGES/LARGE/JF-2018-02444Z_0007.JPEG.

[5] B. Liaset, E. Lied, M. Espe, Enzymatic hydrolysis of by-products from the fish-filleting industry; chemical characterisation and nutritional evaluation, J. Sci. Food Agric. 80 (2000) 581–589, https://doi.org/10.1002/(SICI)1097-0010(200004)80:5.

[6] J.A. Vázquez, A. Meduíña, A.I. Durán, M. Nogueira, A. Fernández-Compás, R. I. Pérez-Martín, I. Rodríguez-Amado, Production of valuable compounds and bioactive metabolites from by-products of fish discards using chemical processing, enzymatic hydrolysis, and bacterial fermentation, Mar. Drugs 17 (2019) 139, https://doi.org/10.3390/MD17030139, 17 (2019) 139.

[7] S.G. Wubshet, D. Lindberg, E. Veiseth-Kent, K.A. Kristoffersen, U. Böcker, K. E. Washburn, N.K. Afseth, Bioanalytical aspects in enzymatic protein hydrolysis of by-products, in: Proteins: Sustainable Source, Processing and Applications, Elsevier, 2019, pp. 225–258, https://doi.org/10.1016/b978-0-12-816695-6.00008-8.

[8] D. Lindberg, K.A. Kristoffersen, H. de Vogel-van den Bosch, S.G. Wubshet, U. Böcker, A. Rieder, E. Fricke, N.K. Afseth, Effects of poultry raw material variation and choice of protease on protein hydrolysate quality, Process Biochem. 110 (2021) 85–93, https://doi.org/10.1016/J.PROCBIO.2021.07.014.

[9] Z. Li, B. Wang, C. Chi, Y. Gong, H. Luo, G. Ding, Influence of average molecular weight on antioxidant and functional properties of cartilage collagen hydrolysates from Sphyrna lewini, Dasyatis akjei and Raja porosa, Food Res. Int. 51 (2013) 283–293, https://doi.org/10.1016/j.foodres.2012.12.031.

[10] V. García Arteaga, M. Apéstegui Guardia, I. Muranyi, P. Eisner, U. Schweiggert-Weisz, Effect of enzymatic hydrolysis on molecular weight distribution, techno-functional properties and sensory perception of pea protein isolates, Innovative Food Sci. Emerging Technol. 65 (2020), 102449, https://doi.org/10.1016/j.ifset.2020.102449.

[11] L. Sorokina, A. Rieder, S. Koga, N.K. Afseth, R.D.C.L. Lima, S.R. Wilson, S. G. Wubshet, Multivariate correlation of infrared fingerprints and molecular weight distributions with bioactivity of poultry by-product protein hydrolysates, J. Funct. Foods 95 (2022), 105170, https://doi.org/10.1016/J.JFF.2022.105170.

[12] R. Šližyte, E. Daukšas, E. Falch, I. Storrø, T. Rustad, Characteristics of protein fractions generated from hydrolysed cod (Gadus morhua) by-products, Process Biochem. 40 (2005) 2021–2033, https://doi.org/10.1016/J.PROCBIO.2004.07.016.

[13] S.G. Wubshet, J.P. Wold, N.K. Afseth, U. Böcker, D. Lindberg, F.N. Ihunegbo, I. Måge, Feed-forward prediction of product qualities in enzymatic protein hydrolysis of poultry by-products: a spectroscopic approach, Food Bioprocess Technol. (2018), https://doi.org/10.1007/s11947-018-2161-y.

[14] S. Grassi, C. Alamprese, Advances in NIR spectroscopy applied to process analytical technology in food industries, Curr. Opin. Food Sci. 22 (2018) 17–21, https://doi.org/10.1016/j.cofs.2017.12.008.

[15] G. Tøgersen, T. Isaksson, B.N. Nilsen, E.A. Bakker, K.I. Hildrum, On-line NIR analysis of fat, water and protein in industrial scale ground meat batches, Meat Sci. 51 (1999) 97–102, https://doi.org/10.1016/S0309-1740(98)00106-5.

[16] J.P. Wold, M. O'Farrel, M. Høy, J. Tschudi, On-line determination and control of fat content in batches of beef trimmings by NIR imaging spectroscopy, Meat Sci. 89 (2011) 317–324.

[17] G. Elmasry, J.P. Wold, High-speed assessment of fat and water content distribution in fish fillets using online imaging spectroscopy, J. Agric. Food Chem. 56 (2008) 7672–7677, https://doi.org/10.1021/JF801074S/ASSET/IMAGES/MEDIUM/JF-2008-01074S_0006.GIF.

[18] J.P. Wold, F. Bjerke, I. Måge, Automatic control of fat content in multiple batches of meat trimmings by process analytical technology, Fleischwirtschaft International 31 (2016) 69–74.

[19] I. Måge, J.P. Wold, F. Bjerke, V. Segtnan, On-line sorting of meat trimmings into targeted fat categories, J. Food Eng. 115 (2013) 306–313, https://doi.org/10.1016/j.jfoodeng.2012.10.030.

[20] O. Monago-Maraña, J.P. Wold, R. Rødbotten, K.R. Dankel, N.K. Afseth, Raman, near-infrared and fluorescence spectroscopy for determination of collagen content in ground meat and poultry by-products, Lebensm. Wiss. Technol. 140 (2021), 110592, https://doi.org/10.1016/J.LWT.2020.110592.

[21] K.A. Esmonde-White, M. Cuellar, C. Uerpmann, B. Lenain, I.R. Lewis, Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing, Anal. Bioanal. Chem. 409 (2017) 637–649, https://doi.org/10.1007/S00216-016-9824-1.

[22] T.A. Lintvedt, P.V. Andersen, N.K. Afseth, K. Heia, S.K. Lindberg, J.P. Wold, Raman spectroscopy and NIR hyperspectral imaging for in-line estimation of fatty acid features in salmon fillets, Talanta 254 (2023), 124113, https://doi.org/10.1016/J.TALANTA.2022.124113.

[23] F. Tao, M. Ngadi, Recent advances in rapid and nondestructive determination of fat content and fatty acids composition of muscle foods, Crit. Rev. Food Sci. Nutr. 58 (2018) 1565–1593, https://doi.org/10.1080/10408398.2016.1261332.

[24] P.V. Andersen, N.K. Afseth, E. Gjerlaug-Enger, J.P. Wold, Prediction of water holding capacity and pH in porcine longissimus lumborum using Raman spectroscopy, Meat Sci. 172 (2021), 108357, https://doi.org/10.1016/J.MEATSCI.2020.108357.

[25] P.V. Andersen, J.P. Wold, N.K. Afseth, Assessment of bulk composition of heterogeneous food matrices using Raman spectroscopy, Appl. Spectrosc. 75 (2021) 1278–1287, https://doi.org/10.1177/00037028211006150.

[26] I. Latka, S. Dochow, C. Krafft, B. Dietzek, J. Popp, Fiber optic probes for linear and nonlinear Raman applications – current trends and future development, Laser Photon. Rev. 7 (2013) 698–731, https://doi.org/10.1002/LPOR.201200049.

[27] H. Wikström, I.R. Lewis, L.S. Taylor, Comparison of sampling techniques for in-line monitoring using Raman spectroscopy, Appl. Spectrosc. 59 (2005) 934–941, https://doi.org/10.1366/0003702054411553.

[28] T.A. Lintvedt, P.v. Andersen, N.K. Afseth, B. Marquardt, L. Gidskehaug, J.P. Wold, Feasibility of in-line Raman spectroscopy for quality assessment in food industry: how fast can we go? Appl. Spectrosc. 76 (2022) 559–568, https://doi.org/10.1177/00037028211056931.

[29] T.A. Lintvedt, P.V. Andersen, N.K. Afseth, J.P. Wold, In-line Raman spectroscopy for characterization of an industrial poultry raw material stream, Talanta 266 (2023), 125079, https://doi.org/10.1016/J.TALANTA.2023.125079.

[30] N.K. Afseth, K. Dankel, P.V. Andersen, G.F. Difford, S.S. Horn, A. Sonesson, B. Hillestad, J.P. Wold, E. Tengstrand, Raman and near infrared spectroscopy for quantification of fatty acids in muscle tissue—a salmon case study, Foods 11 (2022) 962, https://doi.org/10.3390/FOODS11070962/S1.

[31] J.P. Wold, F. Lundby, B. Egelandsdal, Quantification of connective tissue (hydroxyproline) in ground beef by autofluorescence spectroscopy, J. Food Sci. 64 (1999) 377–383, https://doi.org/10.1111/J.1365-2621.1999.TB15045.X.

[32] J.P. Wold, M. Mielnik, Nondestructive assessment of lipid oxidation in minced poultry meat by autofluorescence spectroscopy, J. Food Sci. 65 (2000) 87–95, https://doi.org/10.1111/J.1365-2621.2000.TB15961.X.

[33] G.A. Wagnières, W.M. Star, B.C. Wilson, In vivo fluorescence spectroscopy and imaging for oncological applications, Photochem. Photobiol. 68 (1998) 603–632, https://doi.org/10.1111/J.1751-1097.1998.TB02521.X.

[34] K.A. Kristoffersen, N.K. Afseth, U. Böcker, D. Lindberg, H. de Vogel-van den Bosch, M.L. Ruud, S.G. Wubshet, Average molecular weight, degree of hydrolysis and dry-film FTIR fingerprint of milk protein hydrolysates: intercorrelation and application in process monitoring, Food Chem. 310 (2020), 125800, https://doi.org/10.1016/J.FOODCHEM.2019.125800.

[35] S.M. Rutherfurd, Methodology for determining degree of hydrolysis of proteins in hydrolysates: a review, J. AOAC Int. 93 (2010) 1515–1522, https://doi.org/10.1093/JAOAC/93.5.1515.

[36] S. Chutipongtanate, K. Watcharatanyatip, T. Homvises, K. Jaturongkakul, V. Thongboonkerd, Systematic comparisons of various spectrophotometric and colorimetric methods to measure concentrations of protein, peptide and amino acid: detectable limits, linear dynamic ranges, interferences, practicality and unit costs, Talanta 98 (2012) 123–129, https://doi.org/10.1016/J.TALANTA.2012.06.058.

[37] S.A. Jaywant, H. Singh, K.M. Arif, Sensors and Instruments for Brix Measurement: A Review, Sensors 22 (2022) 2290, https://doi.org/10.3390/S22062290, 22 (2022) 2290.

[38] Y. Zhang, L. Luo, J. Li, S. Li, W. Qu, H. Ma, A.O. Oladejo, X. Ye, In-situ and real-time monitoring of enzymatic process of wheat gluten by miniature fiber NIR spectrometer, Food Res. Int. 99 (2017) 147–154, https://doi.org/10.1016/J.FOODRES.2017.03.048.

[39] Y. Zhang, Y. Li, S. Li, H. Zhang, H. Ma, In situ monitoring of the effect of ultrasound on the sulfhydryl groups and disulfide bonds of wheat gluten, Molecules 23 (2018) 1376, https://doi.org/10.3390/MOLECULES23061376, 23 (2018) 1376.

[40] D. Lindberg, K.A. Kristoffersen, S.G. Wubshet, L.M.G. Hunnes, M. Dalsnes, K. R. Dankel, V. Høst, N.K. Afseth, Exploring effects of protease choice and protease combinations in enzymatic protein hydrolysis of poultry by-products, Molecules 26 (2021) 5280, https://doi.org/10.3390/MOLECULES26175280, 26 (2021) 5280.

[41] M. Gavrilov, M.J. Monteiro, Derivation of the molecular weight distributions from size exclusion chromatography, Eur. Polym. J. 65 (2015) 191–196, https://doi.org/10.1016/J.EURPOLYMJ.2014.11.018.

[42] D.W. Shortt, Differential molecular weight distributions in high performance size exclusion chromatography, J. Liq. Chromatogr. 16 (1993) 3371–3391, https://doi.org/10.1080/10826079308019695.

[43] K. Štulík, V. Pacáková, M. Tichá, Some potentialities and drawbacks of contemporary size-exclusion chromatography, J. Biochem. Biophys. Methods 56 (2003) 1–13, https://doi.org/10.1016/S0165-022X(03)00053-8.

[44] A. Barth, Infrared spectroscopy of proteins, Biochim. Biophys. Acta Bioenerg. 1767 (2007) 1073–1101, https://doi.org/10.1016/J.BBABIO.2007.06.004.

[45] C. Ruckebusch, N. Nedjar-Arroume, S. Magazzeni, J.P. Huvenne, P. Legrand, Hydrolysis of haemoglobin surveyed by infrared spectroscopy: I. solvent effect on the secondary structure of haemoglobin, J. Mol. Struct. 478 (1999) 185–191, https://doi.org/10.1016/S0022-2860(98)00753-4.

[46] N.A. Poulsen, C.E. Eskildsen, M. Akkerman, L.B. Johansen, M.S. Hansen, P. W. Hansen, T. Skov, L.B. Larsen, Predicting hydrolysis of whey protein by mid-infrared spectroscopy, Int. Dairy J. 61 (2016) 44–50, https://doi.org/10.1016/J.IDAIRYJ.2016.04.002.

[47] S.G. Wubshet, I. Måge, U. Böcker, D. Lindberg, S.H. Knutsen, A. Rieder, D. A. Rodriguez, N.K. Afseth, FTIR as a rapid tool for monitoring molecular weight distribution during enzymatic protein hydrolysis of food processing by-products, Anal. Methods 9 (2017) 4247–4254, https://doi.org/10.1039/C7AY00865A.

[48] U. Böcker, S.G. Wubshet, D. Lindberg, N.K. Afseth, Fourier-transform infrared spectroscopy for characterization of protein chain reductions in enzymatic reactions, Analyst 142 (2017) 2812–2818, https://doi.org/10.1039/C7AN00488E.
[49] K.A. Kristoffersen, K.H. Liland, U. Böcker, S.G. Wubshet, D. Lindberg, S.J. Horn, N. K. Afseth, FTIR-based hierarchical modeling for prediction of average molecular weights of protein hydrolysates, Talanta 205 (2019), 120084, https://doi.org/10.1016/j.talanta.2019.06.084.
[50] B. Kafle, U. Bocker, S.G. Wubshet, K. Dankel, I. Mage, M. Farrell, N.K. Afseth, Fourier-transform infrared spectroscopy for characterization of liquid protein solutions: a comparison of two sampling techniques, Vib. Spectrosc. 124 (2023), 103490, https://doi.org/10.1016/J.VIBSPEC.2022.103490.
[51] I. Måge, U. Böcker, S.G. Wubshet, D. Lindberg, N.K. Afseth, Fourier-transform infrared (FTIR) fingerprinting for quality assessment of protein hydrolysates, LWT - Food Sci. Technol. (Lebensmittel-Wissenschaft -Technol.) 152 (2021), 112339, https://doi.org/10.1016/j.lwt.2021.112339.
[52] K.A. Kristoffersen, I. Måge, S.G. Wubshet, U. Böcker, K. Riiser Dankel, A. Lislelid, M.A. Rønningen, N.K. Afseth, FTIR-based prediction of collagen content in hydrolyzed protein samples, Spectrochim. Acta Mol. Biomol. Spectrosc. 301 (2023), 122919, https://doi.org/10.1016/J.SAA.2023.122919.
[53] K.E. Anderssen, E.R. McCarney, Online monitoring of enzymatic hydrolysis of marine by-products using benchtop nuclear magnetic resonance spectroscopy, Food Control 112 (2020), 107053, https://doi.org/10.1016/J.FOODCONT.2019.107053.
[54] U.K. Sundekilde, L. Jarno, N. Eggers, H.C. Bertram, Real-time monitoring of enzyme-assisted animal protein hydrolysis by NMR spectroscopy – an NMR reactomics concept, Lebensm. Wiss. Technol. 95 (2018) 9–16, https://doi.org/10.1016/j.lwt.2018.04.055.
[55] S. Steinsholm, Å. Oterhals, J. Underhaug, I. Måge, A. Malmendal, T. Aspevik, Sensory assessment of fish and chicken protein hydrolysates. Evaluation of NMR metabolomics profiling as a new prediction tool, Cite This: J. Agric. Food Chem. 68 (2020) 3890, https://doi.org/10.1021/acs.jafc.9b07828.
[56] L.A. Colnago, F.D. Andrade, A.A. Souza, R.B.V. Azeredo, A.A. Lima, L.M. Cerioni, T. M. Osán, D.J. Pusiol, Why is inline NMR rarely used as industrial sensor? Challenges and opportunities, Chem. Eng. Technol. 37 (2014) 191–203, https://doi.org/10.1002/CEAT.201300380.
[57] I. Bøgwald, T.K.K. Østbye, A.M. Pedersen, S.B. Rønning, J. Dias, K.E. Eilertsen, S. G. Wubshet, Calanus finmarchicus hydrolysate improves growth performance in feeding trial with European sea bass juveniles and increases skeletal muscle growth in cell studies, Sci. Rep. 13 (1) (2023) 1–14, https://doi.org/10.1038/s41598-023-38970-5.
[58] M.A. Amiza, S. Nurul Asikin, A.L. Faazaz, Optimization of enzymatic protein hydrolysis from silver catfish, Int. Food Res. J. 18 (2011) 775–781.
[59] K.A. Kristoffersen, N.K. Afseth, U. Böcker, K.R. Dankel, M.A. Rønningen, A. Lislelid, R. Ofstad, D. Lindberg, S.G. Wubshet, Post-enzymatic hydrolysis heat treatment as an essential unit operation for collagen solubilization from poultry by-products, Food Chem. 382 (2022), https://doi.org/10.1016/j.foodchem.2022.132201.
[60] R.S. Kenett, G. Shmueli, On information quality, J R Stat Soc Ser A Stat Soc 177 (2014) 3–38, https://doi.org/10.1111/RSSA.12007.
[61] M. Thiel, B. Féraud, B. Govaerts, ASCA+ and APCA+: extensions of ASCA and APCA in the analysis of unbalanced multifactorial designs, J. Chemom. 31 (2017) e2895, https://doi.org/10.1002/cem.2895.
[62] J.J. Jansen, H.C.J. Hoefsloot, J. van der Greef, M. Timmerman, J.A. Westerhuis, A. K. Smilde, ASCA: analysis of multivariate data obtained from an experimental design, J. Chemom. 19 (2005) 469–481, https://doi.org/10.1002/cem.952.
[63] K.H. Liland, A. Smilde, F. Marini, T. Naes, Confidence ellipsoids for ASCA models based on multivariate regression theory, J. Chemom. 32 (2018), https://doi.org/10.1002/cem.2990.
[64] M. Martin, B. Govaerts, LiMM-PCA: combining ASCA+ and linear mixed models to analyse high-dimensional designed data, J. Chemom. (2020), https://doi.org/10.1002/cem.3232.
[65] M. de Figueiredo, S. Giannoukos, S. Rudaz, R. Zenobi, J. Boccard, Efficiently handling high-dimensional data from multifactorial designs with unequal group sizes using Rebalanced ASCA (RASCA), J. Chemom. (2022), e3401, https://doi.org/10.1002/CEM.3401.
[66] P.D.B. Harrington, N.E. Vieira, J. Espinoza, J.K. Nien, R. Romero, A.L. Yergey, Analysis of variance–principal component analysis: a soft tool for proteomic discovery, Anal. Chim. Acta 544 (2005) 118–127, https://doi.org/10.1016/J.ACA.2005.02.042.
[67] D. Jouan-Rimbaud Bouveresse, R.C. Pinto, L.M. Schmidtke, N. Locquet, D. N. Rutledge, Identification of significant factors by an extension of ANOVA–PCA based on multi-block analysis, Chemometr. Intell. Lab. Syst. 106 (2011) 173–182, https://doi.org/10.1016/J.CHEMOLAB.2010.05.005.
[68] Ø. Langsrud, 50–50 multivariate analysis of variance for collinear responses, J. Roy. Stat. Soc. D. 51 (2002) 305–317. http://onlinelibrary.wiley.com/doi/10.1111/1467-9884.00320/full. (Accessed 30 September 2014).
[69] A. el Ghaziri, E.M. Qannari, T. Moyon, M.-C. Alexandre-Gouabau, AoV-PLS: a new method for the analysis of multivariate data depending on several factors, Electronic Journal of Applied Statistical Analysis 8 (2015) 214–235, https://doi.org/10.1285/i20705948v8n2p214.
[70] F. Marini, D. de Beer, E. Joubert, B. Walczak, Analysis of variance of designed chromatographic data sets: the analysis of variance-target projection approach, J. Chromatogr. A 1405 (2015) 94–102, https://doi.org/10.1016/J.CHROMA.2015.05.060.
[71] A.K. Smilde, T. Næs, K.H. Liland, Multiblock Data Fusion in Statistics and Machine Learning: Applications in the Natural and Life Sciences, John Wiley & Sons, 2022.
[72] T. Næs, O. Tomic, N.K. Afseth, V. Segtnan, I. Måge, Multi-block regression based on combinations of orthogonalisation, PLS-regression and canonical correlation analysis, Chemometr. Intell. Lab. Syst. 124 (2013) 32–42, https://doi.org/10.1016/j.chemolab.2013.03.006.
[73] A. Biancolillo, I. Måge, T. Næs, Combining SO-PLS and linear discriminant analysis for multi-block classification, Chemometr. Intell. Lab. Syst. 141 (2015) 58–67, https://doi.org/10.1016/j.chemolab.2014.12.001.
[74] A. Biancolillo, T. Næs, R. Bro, I. Måge, Extension of SO-PLS to multi-way arrays: SO-N-PLS, Chemometr. Intell. Lab. Syst. 164 (2017) 113–126, https://doi.org/10.1016/j.chemolab.2017.03.002.
[75] A. Biancolillo, K.H. Liland, I. Måge, T. Næs, R. Bro, Variable selection in multi-block regression, Chemometr. Intell. Lab. Syst. 156 (2016) 89–101, https://doi.org/10.1016/j.chemolab.2016.05.016.
[76] J. Roger, A. Biancolillo, F. Marini, Sequential preprocessing through ORThogonalization (SPORT) and its application to near infrared spectroscopy, Chemometr. Intell. Lab. Syst. 199 (2020), 103975, https://doi.org/10.1016/j.chemolab.2020.103975.
[77] F.A.A. Souza, R. Araújo, J. Mendes, Review of soft sensor methods for regression applications, Chemometr. Intell. Lab. Syst. 152 (2016) 69–79, https://doi.org/10.1016/J.CHEMOLAB.2015.12.011.
[78] Y. Jiang, S. Yin, J. Dong, O. Kaynak, A review on soft sensors for monitoring, control, and optimization of industrial processes, IEEE Sensor. J. 21 (2021) 12868–12881, https://doi.org/10.1109/JSEN.2020.3033153.
[79] F. Capaci, E. Vanhatalo, M. Kulahci, B. Bergquist, The revised Tennessee Eastman process simulator as testbed for SPC and DoE methods, Qual. Eng. 31 (2019) 212–229, https://doi.org/10.1080/08982112.2018.1461905.
[80] T. Offermans, E. Szymańska, L.M.C. Buydens, J.J. Jansen, Synchronizing process variables in time for industrial process monitoring and control, Comput. Chem. Eng. 140 (2020), 106938, https://doi.org/10.1016/j.compchemeng.2020.106938.
[81] Q. Jiang, X. Yan, B. Huang, Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes, Ind. Eng. Chem. Res. 58 (2019) 12899–12912, https://doi.org/10.1021/acs.iecr.9b02391.
[82] M.S. Reis, G. Gins, Industrial process monitoring in the big data/industry 4.0 era: from detection, to diagnosis, to prognosis, Processes 5 (2017), https://doi.org/10.3390/pr5030035.
[83] T. Offermans, T.H. Wijker, R. Folcarelli, R. Heemskerk, P.P. Lamers, M. Proença, T. N. Tran, L.M.C. Buydens, J.J. Jansen, ENDBOSS, Industrial endpoint detection using batch-specific control spaces of spectroscopic data, Chemometr. Intell. Lab. Syst. 209 (2021), 104229, https://doi.org/10.1016/J.CHEMOLAB.2020.104229.
[84] F. Westad, L. Gidskehaug, B. Swarbrick, G.R. Flåten, Assumption free modeling and monitoring of batch processes, Chemometr. Intell. Lab. Syst. 149 (2015) 66–72, https://doi.org/10.1016/j.chemolab.2015.08.022.
[85] M. Kuchta, S.G. Wubshet, N.K. Afseth, K.A. Mardal, K.H. Liland, Encoder–decoder neural networks for predicting future FTIR spectra – application to enzymatic protein hydrolysis, J. Biophot. 15 (2022), https://doi.org/10.1002/jbio.202200097.
[86] R.S. Rao, C.G. Kumar, R.S. Prakasham, P.J. Hobbs, The Taguchi methodology as a statistical tool for biotechnological applications: a critical appraisal, Biotechnol. J. 3 (2008) 510–523, https://doi.org/10.1002/BIOT.200700201.
[87] I. Berget, T. Næs, Optimal sorting of raw materials, based on the predicted end-product quality, Qual. Eng. 14 (2002) 459–478, https://doi.org/10.1081/QEN-120001883.
[88] I. Berget, A. Aamodt, E. Mosleth Færgestad, T. Næs, Optimal sorting of raw materials for use in different products, Chemometr. Intell. Lab. Syst. 67 (2003) 79–93, https://doi.org/10.1016/S0169-7439(03)00052-2.
[89] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B.D. Youn, M.D. Todd, S. Mahadevan, C. Hu, Z. Hu, A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies, Struct. Multidiscip. Optim. 65 (2022) 1–55, https://doi.org/10.1007/S00158-022-03425-4/FIGURES/28.
[90] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B.D. Youn, M.D. Todd, S. Mahadevan, C. Hu, Z. Hu, A comprehensive review of digital twin—part 2: roles of uncertainty quantification and optimization, a battery digital twin, and perspectives, Struct. Multidiscip. Optim. 66 (2023) 1–43, https://doi.org/10.1007/S00158-022-03410-X/TABLES/4.
[91] M. Sokolov, M. von Stosch, H. Narayanan, F. Feidl, A. Butté, Hybrid modeling — a key enabler towards realizing digital twins in biopharma? Curr Opin Chem Eng 34 (2021), 100715 https://doi.org/10.1016/J.COCHE.2021.100715.