

Provided by the author(s) and NUI Galway in accordance with publisher policies. Please cite the published version when available.

|                             |  |
|-----------------------------|--|
| Title                       | Prediction of Cell Culture Media performance using Fluorescence Spectroscopy   |
| Author(s)                   | Ryan, Paul W.; Li, Boyan; Shanahan, Michael; Ryder, Alan G.  |
| Publication Date            | 2010   |
| Publication Information     | Ryan, P.W., Li, B., Shanahan, M., Leister, K.J., Ryder, A.G. (2010) 'Prediction of Cell Culture Media performance using Fluorescence Spectroscopy'. <i>Analytical Chemistry</i> , 82 :1311-1317.           |
| Publisher                   | ACS Publications   |
| Link to publisher's version | <a href="http://dx.doi.org/10.1021/ac902337c">http://dx.doi.org/10.1021/ac902337c</a>  |
| Item record                 | <a href="http://www.nuigalway.ie/nanoscale/publications.html">http://www.nuigalway.ie/nanoscale/publications.html</a> ;<br><a href="http://hdl.handle.net/10379/2932">http://hdl.handle.net/10379/2932</a> |
| DOI                         | <a href="http://dx.doi.org/http://dx.doi.org/10.1021/ac902337c">http://dx.doi.org/http://dx.doi.org/10.1021/ac902337c</a>  |

Downloaded 2020-10-23T09:15:47Z

Some rights reserved. For more information, please see the item record link above.



# PREDICTION OF CELL CULTURE MEDIA PERFORMANCE USING FLUORESCENCE SPECTROSCOPY.

Paul W. Ryan,<sup>1,2</sup> Boyan Li,<sup>1,2</sup> Michael Shanahan,<sup>1,2</sup> Kirk J. Leister,<sup>3</sup> and Alan G. Ryder.<sup>1,2\*</sup>

<sup>1</sup> Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Galway, Ireland.

<sup>2</sup> Centre for Bioanalytical Sciences, School of Chemistry, National University of Ireland, Galway, Galway, Ireland.

<sup>3</sup> Bristol-Myers Squibb, Process Analytical Sciences, Syracuse, New York, USA.

\* Corresponding author: **Email:** alan.ryder@nuigalway.ie **Phone:** +353-91-492943.

**Published Citation:** Prediction of Cell Culture Media performance using Fluorescence Spectroscopy. P.W. Ryan, B. Li, M. Shanahan, K.J. Leister, and A.G. Ryder. *Analytical Chemistry*, **82**(4), 1311-1317, (2010).

**DOI:** <http://dx.doi.org/10.1021/ac902337c>

**Note:** This was the final revised version as accepted, this version does not include any proofing error correction. The definitive, final version of the paper is available on the Analytical Chemistry website.

## Abstract:

Cell culture media used in industrial mammalian cell culture are complex aqueous solutions that are inherently difficult to analyze comprehensively. The analysis of media quality and variance is of utmost importance in efficient manufacturing. We are exploring the use of rapid “holistic” analytical methods that can be used for routine screening of cell culture media used in industrial biotechnology. The application of rapid fluorescence spectroscopic techniques to the routine analysis of cell culture media (Chinese Hamster Ovary cell based manufacture) was investigated. We have developed robust methods which can be used to identify compositional changes and ultimately predict the efficacy of individual fed batch media in terms of downstream protein product yield with an accuracy of  $\pm 0.13$  g/L. This is achieved through the implementation of chemometric methods such as multiway robust principal component analysis

(MROBPCA), n-way partial least squares-discriminant analysis and regression (NPLS-DA and NPLS). This ability to observe compositional changes and predict product yield before media use has enormous potential and should permit the effective elimination of one of the major process variables leading to more consistent product quality and improved yield. These robust and reliable methods have the potential to become an important part of upstream biopharmaceutical quality control and analysis.

**Keywords:** Fluorescence, cell culture, media, chemometrics.

## INTRODUCTION

The essential purpose of cell culture media is to produce the correct physiological environment to allow effective *in vitro* culturing of mammalian cells by providing survival and growth requirements. Since the physiological environments of most cells are not fully defined, this is a difficult task and formulation of appropriate media requires the careful selection and blending of a wide variety of components. The resulting media are highly complex mixtures, containing a variety of amino acids, carbohydrates, cofactors, and other materials. These aqueous based media must provide all of the cell nutrients required for growth as well as providing an energy source, while maintaining pH and osmolarity.

In mammalian cell fermentation there are two designated types of cell culture media used in the fed-batch process: basal media and feed media. Examples of basal media include Dulbeccos' Modified Eagles Medium (DMEM), Roswell Park Memorial Institute (RPMI) Medium, Iscove modified Dulbeccos' medium and Ham's Nutrient Mixture medium (F12).<sup>1,2</sup> Basal media plus important growth factors are added at the initial bioreactor inoculation step and provide specific cell nutrients. Cell proliferation occurs in the basal media until the desired population density is achieved. At this point feed media is added to promote the bioreactor cell mass, a production phase of the manufacturing process whereby the cells shift energy to protein synthesis. This production phase is where the genetically engineered biotherapeutic protein product is made. Feed media are used in fed-batch processes and media supply is strictly controlled to optimise growth rate, product yield and quality, while preventing the formation of unwanted metabolites. Feed media are a complex, nutrient enriched formulation of inorganic salts, sugars, amino acids, vitamins, metals, and other nutrients (e.g., growth factors, yeast/soy extracts, etc.). The composition of the feed media and feed rate vary according to cell type and product being produced.<sup>2,3,4</sup>

As industrial cell culture becomes more widespread so too does the demand for rapid, reliable, robust and non-destructive analytical methods that can be integrated into process control to improve end product yield and quality. The complexity of these mammalian cell culture media means that efficient analysis, using current separations based analytical methods, is challenging, time consuming, and expensive.<sup>1,5,6</sup> Therefore, the development of analytical methods with a rapid holistic approach to sample evaluation is more appropriate than attempting to determine every individual component (which is not feasible).

Fluorescence spectroscopy is widely used for the analysis of biogenic materials and thus is a useful option for the holistic analysis of complex materials.<sup>7</sup> Unfortunately, conventional single excitation wavelength spectroscopy does not yield sufficiently detailed spectra suitable for complex sample characterization. However, it is possible with most scanning fluorimeters to rapidly record both excitation and emission spectra, yielding an excitation-emission matrix (EEM) spectrum. The tri-parametric EEM data incorporates information on excitation wavelength, emission wavelength, and fluorescence intensity. With EEM, the spectral information content is increased making it easier to accurately characterize multicomponent samples. Because the technique is non-contact, non-destructive, and optically based, it is ideal for on-line, *in situ* and *in vivo* measurements via fiber optical systems. Other advantages include: high sensitivity, high signal to noise ratios, and relatively large linear ranges. EEM and other multidimensional fluorescence (MDF) spectroscopic methods have found wide application,<sup>8-14</sup> from dye analysis in heritage science,<sup>15</sup> to crude oil characterisation,<sup>16</sup> to beer process analysis,<sup>17,18</sup> to human tissue analysis,<sup>19,20</sup> and the characterization of dissolved organic matter (DOM) in water.<sup>21-26</sup> The application of MDF methods to online process control in the context of monitoring metabolic changes and turbidity has also been implemented.<sup>27</sup>

Since EEM spectra cover a wide range of excitation and emission wavelengths, it allows for the simultaneous detection of a wide variety of biogenic fluorophores present in complex, turbid, media environments such as amino acids, proteins, coenzymes, and vitamins. To extract quantitative and qualitative data from EEM spectra, advanced mathematical treatment and analysis of the information rich spectra is required. Chemometric analysis of EEM data is the key factor in developing robust analytical methods. Methods such as partial least squares regression (PLS) and parallel factor analysis (PARAFAC) have been used for quantitative determination of various pharmaceuticals during formulation<sup>28</sup> as well as in biological fluids,<sup>29,30</sup> and also for the characterization of DOM.<sup>31,32</sup>

Our group has been focusing on developing innovative optical spectroscopy based analytical methods for the rapid analysis of complex cell culture media. By combining fluorescence methods in conjunction with Raman spectroscopy, the majority of all commonly used materials throughout an entire biopharmaceutical process can be monitored both rapidly and effectively. Here we demonstrate the use of EEM and chemometrics to rapidly assess the quality of blended cell culture media, discriminate different blend types, observe storage induced changes, and correlate spectral data with product yield. This ability to predict whether a feed media is “good” or “poor” (based on product yield) before addition to a bioreactor is a powerful application of the developed techniques and should result in major cost savings for the industry by means of greatly improved protein product yield consistency.

## **MATERIALS AND METHODS**

**Cell culture media samples:** The feed media lots were manufactured for two different mammalian cell culture processes (A- and L-processes), on a large scale (5,000 L+) by Bristol-Myers Squibb (Syracuse, NY). Each lot was manufactured in a single batch and stored for 20+ days at 2-8 °C and used as required during the bioprocess. Feed media were blended with appropriate nutrient and growth factors at approximately T=0. Aliquots from 33 different lots were removed for analysis at four different timepoints over the 20 day storage period (Table 1). The exact composition of the media is a commercial trade secret and as such was unknown to the analysts in Ireland. All media were sampled and aliquotted under sterile conditions and shipped to Ireland at low temperature, under monitored conditions. These samples were then aliquotted to smaller volumes and stored frozen at -70°C in order to limit further changes in media composition. For analysis, the samples were randomly removed from cold storage and defrosted at room temperature. Once defrosted the samples were handled under aseptic conditions and diluted (50 µL in 1 mL) with ultrapure water before being pipetted into cuvettes and sealed, all measurements were made within 6 hours of defrosting.

**Instrumentation and data collection:** Fluorescence spectra were measured at 25 °C using a Cary Eclipse Fluorescence Spectrometer fitted with a thermostatted (Peltier) 4-position multi-cell holder (Varian). EEM data were collected with spectral ranges of 230 to 520 nm for excitation and 270 to 600 nm emission, a data interval of 5 nm, and measurement typically took ~7 minutes. Semi micro quartz cuvettes (Lightpath Optical Ltd., UK) with a long path length of 10 mm (excitation axis) and a short path length of 2 mm (emission axis) were employed for all measurements. A water background spectrum was measured at the beginning and end of each sample collection run. Three EEM spectra were collected for every sample on different random dates over a 6-month period.

**Data pre-processing and analysis:** Rayleigh and Raman scatter are largely unrelated to the chemical properties of the sample and the scatter peaks do not behave linearly (or trilinearly). This may complicate and bias fluorescence data modelling,<sup>33</sup> and so Rayleigh and Raman bands were removed prior to analysis. The bands were removed by replacement with a linear fit, connecting points either side of the peak using imputation.<sup>34,35</sup> All calculations were performed using MATLAB<sup>®</sup> (ver. 7.4) on a standard PC.<sup>37</sup> For the majority of the analyses undertaken, PLS\_Toolbox 4.0<sup>®</sup>, supplemented by in-house-written codes was used.<sup>36,37</sup>

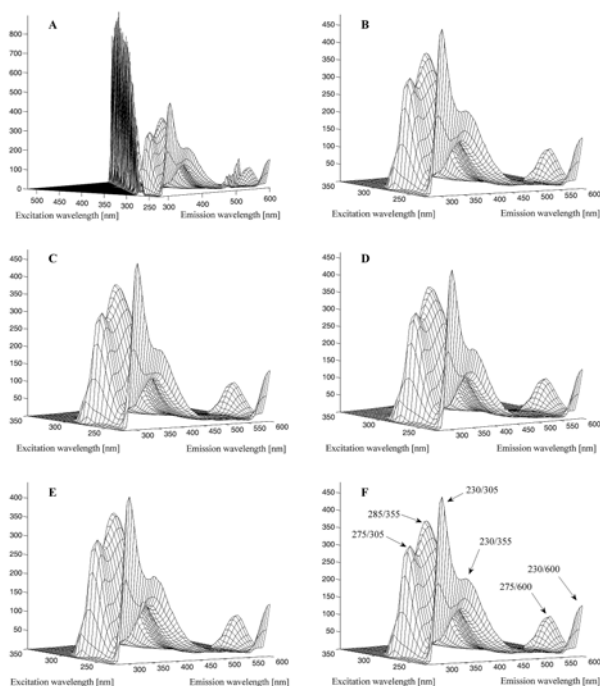
**Methodology:** A range of chemometric methods have to be employed for the robust quantitative analysis of cell culture media. First the spectral quality and sample outliers must be assessed and identified using multiway robust principal component analysis (MROBPCA). Second, changes in media composition with respect to storage time are evaluated using n-way partial least squares-discriminant analysis (NPLS-DA). Finally the quality of the cell culture media is measured by using n-way partial least squares (NPLS) to correlate with protein product yield.

Brief descriptions of the chemometric methods employed, and the requisite references are provided in the supplemental information.<sup>38-58</sup>

## RESULTS AND DISCUSSION

**EEM Spectra:** The media samples are reasonably fluorescent (Figure 1) when illuminated by UV-blue light, with the Rayleigh scattering band being about twice the intensity of the maximum fluorescence signal. The complexity of the EEM spectra of all samples (Figure 1) indicates that fluorescence originates from a variety of fluorophores, the most significant of which are tryptophan and tyrosine, which are present in appreciable quantities: ~1.0 and ~4.8 mg/L respectively. The interplay of energy transfer (ET) and quenching between all of the various components produces a unique EEM profile which can be used to characterize these complex media. Small variances in media composition can cause changes in tryptophan or tyrosine (or both) emission, causing changes in the spectral profile which can be observed. Knowing the exact identity (or concentrations) of the individual absorbing and emitting species is not required (or possible in this case) for the use of EEM data for media characterization.

There are six distinct local maxima within these EEM landscapes and Table 1 gives details of the mean intensity values of the maxima as a function of storage time (Figure 1F). While it is difficult to assign these peaks only to specific fluorophores due to the complex nature of the samples, peaks with 305 nm emission probably originate from tyrosine while emission peaks at 355 nm are likely to result from tryptophan. The emission peaks at 600 nm are second order tyrosine emission (validated by experiment, data not shown). While there are no significant spectral changes evident in the samples with storage time, there does appear to be a slight reduction in the intensity of the tyrosine bands as a function of storage time which is accompanied by a slight increase in the intensity of peaks with emission at 355 nm. This trend would appear to correlate well with a change in ET efficiency from tyrosine to tryptophan with storage time due to compositional changes of the media. It is also worth noting that the majority of this shift in intensity occurs within the first five days of media storage.



**Figure 1:** EEM landscapes of representative samples of feed medias#1-4. (A) Original Media#1 spectrum. Rayleigh corrected spectra of Medias: #1 (B), #2 (C), #3 (D), and #4 (E). (F) Rayleigh corrected Media#1 sample showing local intensity maxima, arrows point to local excitation/emission peaks (Table 1).

**Table 1:** Summary of sample types analyzed and mean fluorescence intensity values of normal samples at the six local maxima as a function of storage time. All the EEM measurements were analyzed by MROBPCA and the outliers omitted prior to calculation of the mean intensity values. A-P = A-process samples, L-P = L-process samples.

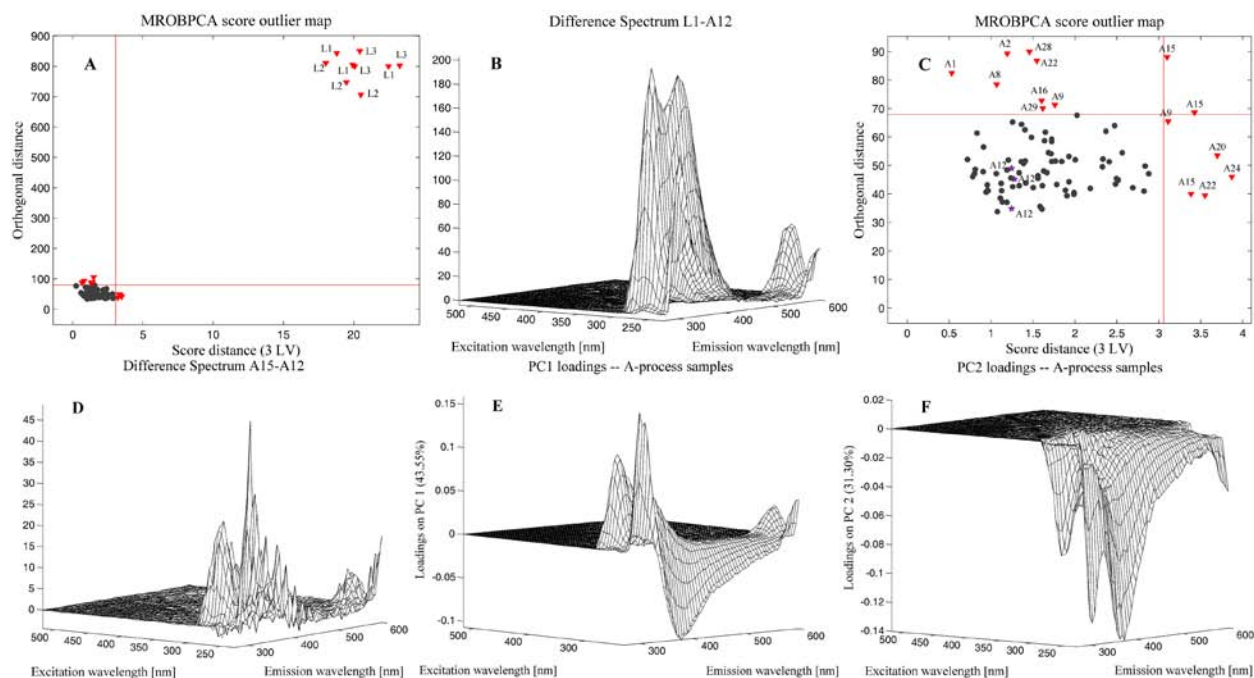
| Material Origin | Storage time (days) | Qty (A-P) | Qty (L-P) | Total Measurements * | Excitation/Emission (nm/nm) |          |          |          |          |          |
|-----------------|---------------------|-----------|-----------|----------------------|-----------------------------|----------|----------|----------|----------|----------|
|                 |                     |           |           |                      | 230 /305                    | 275 /305 | 230 /600 | 275 /600 | 230 /355 | 285 /355 |
| Media#1         | 0                   | 30        | 3         | 75*                  | 481.5                       | 334.2    | 148.4    | 101.1    | 222.4    | 385.5    |
| Media#2         | 5                   | 30        | 4         | 77*                  | 465.8                       | 321.1    | 143.2    | 97.2     | 224.7    | 390.5    |
| Media#3         | 10                  | 27        | 4         | 66*                  | 460.2                       | 322.3    | 142.4    | 97.3     | 221.4    | 388.7    |
| Media#4         | 20+                 | 29        | 4         | 72*                  | 460.9                       | 317.5    | 141.6    | 95.8     | 223.9    | 390.2    |

\* Outliers removed from each sample set.

**MROBPCA media evaluation:** The first step in media analysis is the evaluation of variance for each sample class and the identification of different blends/batches, in essence, a rapid assessment of how similar and reproducible the spectra and samples are. For the 33 Media#1 samples, the 99 spectra after Rayleigh and Raman scattering removal, were analyzed by MROBPCA using three robust PCs which identified a total of 24 outlier measurements from 14 different samples. The nine most significant outliers (Figure 2A) were the L-process samples (L1-L3) which originated from a different CHO based production process to the other samples and were of a slightly different chemical composition to the A-process samples. Figure 2A clearly shows that these disparate media are easily discriminated using EEM and MROBPCA. Figure 2B shows the difference spectrum between typical A-process and L-process samples and it's clear that the major differences are in the tyrosine/tryptophan emission. The L-process samples also have more intense EEM spectra but as yet we cannot ascribe what causes this (increased fluorescence or reduced quenching).

The remaining 15 outlying measurements originated from 11 different A-process samples and the majority of these were singular events in that the remaining two measurements from the triplicate data were within the confidence thresholds. Thus these are most likely due to instrumental/measurement/operator events. The exceptions to this are samples A9, A22, and A15 which gave two or more outlying measurements (Figure 2C). This behavior suggests that significant compositional changes have occurred in these samples which results in their discrimination. Comparing A15 to a regular sample A12 (Figure 2D) by means of their difference spectrum (A15–A12), one observes that the intensity of peaks at Ex/Em 230/305 nm and 275/305 nm are increased in sample A15 suggesting an increase in tyrosine emission. The loadings plots (Figure 2E-F) reveal the range of variation present within the Media#1A-process sample set. The PC1 loadings suggest that the majority of outliers show increased tyrosine/reduced tryptophan emissions as the plot shows an increase in emission intensity at 305 nm and 600 nm coupled with a decrease in intensity at 355 nm.





**Figure 2:** (A) MROBPCA robust score outlier map for 33 Media#1 samples, which classifies the L-Process outliers by defining two threshold values of Score Distance (SD)=3.06 and Orthogonal Distance (OD) = 79.14 (solid straight lines); (B) Difference spectrum between L1 and A12 samples; (C) MROBPCA outlier map for A-process samples, the solid straight lines define threshold values of SD=3.06 and OD=67.93, (D) Difference spectrum between A15 and A12 samples; and MROBPCA loadings plots for PC1 (E) and PC2 (F) obtained from 90 EEM measurements of 30 Media#1 A-process samples after being Rayleigh scattering corrected. The first three robust PCs and a confidence level of 95% were employed for all MROBPCA analyses.

This is in good agreement with the difference spectrum (Figure 2D) which shows the difference between a regular A-process sample and the largest outlier from this set of samples (A15). The PC2 loadings appear to represent those outlying measurements which gave lower intensity signals through the entire spectral range while the loadings on PC3 (see supplemental information) may represent those samples which present reduced tryptophan emission as the plot reveals a reduction in intensity at Ex/Em 285/355 nm.

Multivariate analysis of variance (MANOVA)<sup>59,60</sup> methods were used to calculate the relative within-group variance ( $RMS_w$ ) for the Media#1 samples (Table 2). The  $RMS_w$  value obtained from all 99 spectra is 2.02%, while that from the 75 normal measurements excluding outliers is 0.11%, indicating that the normal measurements are all very similar. Repeating the MROBPCA evaluation and MANOVA analysis for the other feed media showed similar trends in  $RMS_w$  values (Table 2). It is also interesting to note that the  $RMS_w$  value from the normal measurements increases marginally as a function of storage time which may indicate that some minor compositional changes occur during storage.

In total, 103 measurement outliers were identified from 393 triplicate measurements of which 45 originated from the L-process samples. Since these L-process samples were formulated slightly differently their discrimination is unsurprising. The 58 outlying measurements were classified into two groups: Group 1 outliers are those sample measurements which proved to be repeatedly outlying (2 from 3 or 3 from 3), and were adjudged to result from real compositional changes, and Group 2 outliers consisted of samples which had 1 out of 3 replicate measurements was anomalous and as such are most due to instrumental or experimental effects (Table 2, & supplemental information). A similar trend is evident for all of the media sample classes with the number of ‘true’ group 1 compositional outliers typically being in the 7-10% range and the number of group 2 instrumental outliers being in the range of 4-10%.

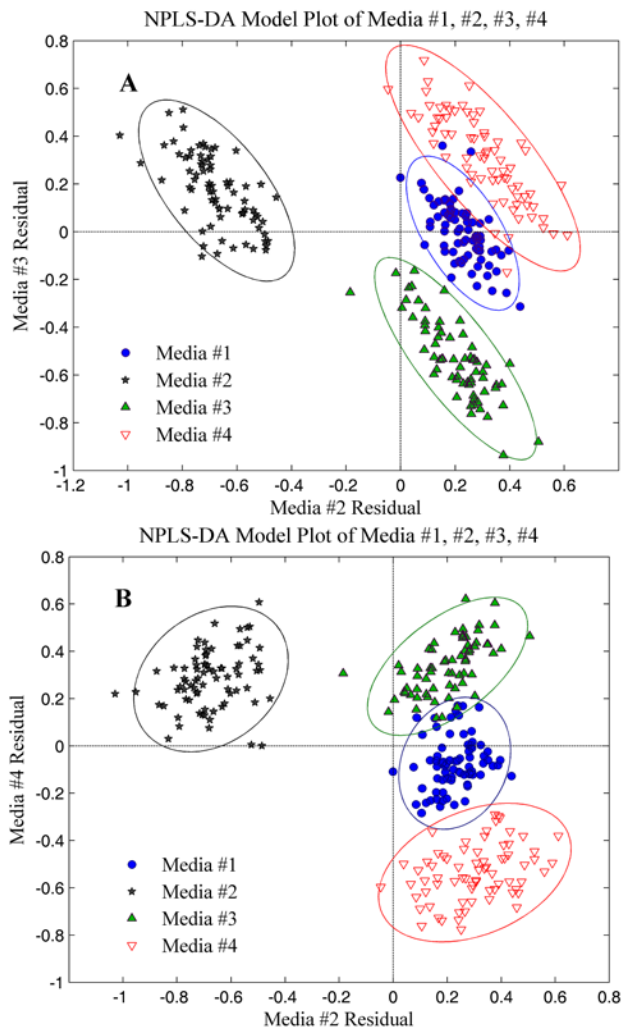
**Table 2:** Overview of MROBPCA outlier evaluation for all Media sample classes. Group 1 = repeatedly outlying measurements. Group 2 = singular outlying measurements. Note that  ${}^bRMS_w$  represents  $RMS_w$  calculations from all measurements in the class group (including outliers) while  ${}^aRMS_w$  represents  $RMS_w$  calculations from the normal measurements only (outliers removed).

| Material Origin | Total Measurements | Outliers  |         |         | Normal Measurements | ${}^bRMS_w(\%)$ | ${}^aRMS_w(\%)$ |
|-----------------|--------------------|-----------|---------|---------|---------------------|-----------------|-----------------|
|                 |                    | L-Process | Group 1 | Group 2 |                     |                 |                 |
| Media#1         | 99                 | 9         | 7       | 8       | 75                  | 2.02            | 0.11            |
| Media#2         | 102                | 12        | 9       | 4       | 77                  | 2.65            | 0.14            |
| Media#3         | 93                 | 12        | 8       | 7       | 66                  | 2.67            | 0.15            |
| Media#4         | 99                 | 12        | 6       | 9       | 72                  | 2.65            | 0.16            |
| Average (%)     |                    | 11.5%     | 7.7%    | 7.4%    | 74%(83.3%*)         | 2.50            | 0.14            |

\* For the A process samples only.

**Storage induced changes in cell culture media:** In some industrial processes, the feed media is manufactured as a single lot, stored under sterile conditions at 2-8 °C and then fed to the reactor at specific timepoints and quantities. From a fundamental chemical point of view, this leads to the possibility of slow rate chemical reactions causing changes in the media composition over time. The ability to detect significant or, subtle compositional changes would be of undoubted interest for quality control and process monitoring. To that end we took all the A-process sample EEM data and used a variety of multiway methods to search for, and identify any spectral changes due to storage effects. From the original data set (393 measurements), the 45 L-process sample, and 58 other outlier measurements were removed before performing MPCA, MROBPCA and SIMCA analyses. All of these methods failed to clearly discriminate samples according to storage time indicating that any compositional changes that could impact on the fluorescence emission, by quenching or ET processes are relatively minor (it does not however rule out changes in components such as glucose which might not be expected to influence fluorescence emission). However, a ROBPCA Raman study on these medias (supplemental information) does not show any large compositional changes.

However, when interpreting the data using NPLS-DA it was possible to correlate changes in residual content with time, indicating that the samples *were* different at the different timepoints. NPLS-DA was performed on the 290 ‘normal’ measurements to generate a model using five latent variables (LVs). It was found that the media timepoints 1-4 could be distinguished from one another by means of their residual information (Figure 3). NPLS-DA fits the spectral data X to numerous dummy Y variables in order to determine which of these most accurately correlates with the data from a number of predefined classes, in this case Media#1, #2, #3, #4. These four Y variable models are then used to predict which samples are within the class associated with the individual models. In typical cases where the samples have significant differences the prediction is quite effective and clear separation is achieved. In our case however the sample differences are minute and hence the prediction was not capable of clear differentiation which led us to utilizing the residual content from the individual Y variable models for classification. We found that the residual content was considerably lower for those samples which were of the class associated with the relevant Y variable model. This indicates that there are subtle changes induced in media composition on extended storage at 2-8 °C.



**Figure 3:** Overview of the NPLS-DA model (five LVs used on the Rayleigh scattering corrected fluorescence EEM data) on the 290 measurements of four types of media samples (Table 1). A solid ellipse specifies a distribution of normal measurements for each class. The plots show that the media composition (as measured by the EEM spectra) is changing subtly on prolonged storage. [More detail in supplementary information, figure S-8]

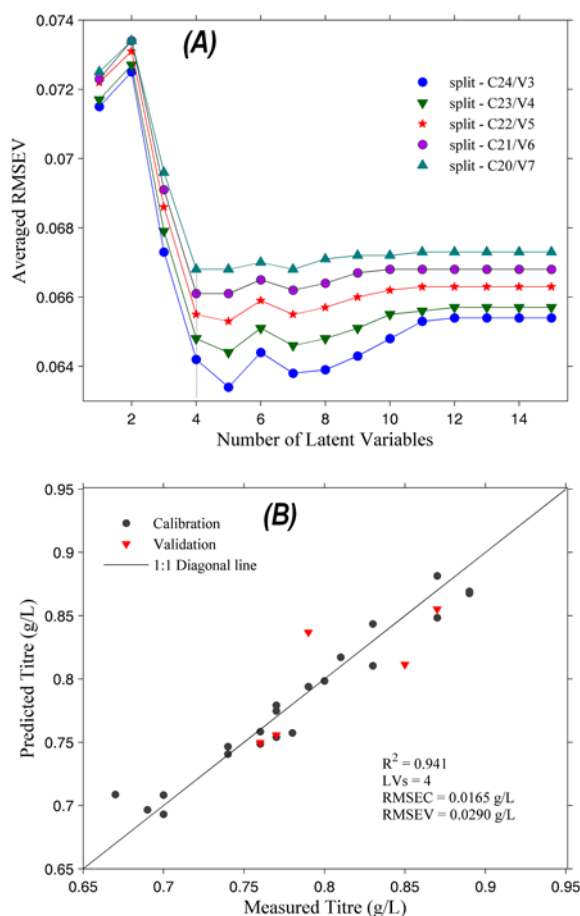
This is a promising finding as it should permit more effective, inexpensive monitoring and thus better control over the relative quality of fed batch media by the identification of potentially detrimental compositional changes prior to the introduction of feed media to the bioreactor. The compositional changes in these medias with storage time were identifiable despite their very small magnitude which gives an indication of the sensitivity level provided by these fluorescence measurements. These minor compositional changes could potentially have a significant effect on the efficacy of the media due to the complex nature of the system and the multitude of roles the media plays in driving the production of the genetically engineered protein products. It follows therefore that any significant compositional change could also be easily identified as demonstrated by the obvious discrimination of L-process from A-process samples.

**Correlation of feed media EEM spectra with protein yield:** The variation evident in the EEM spectra from samples taken at the same timepoint indicates that there are differences between the different lots of feed media. Understanding if these differences are significant is important as it could provide a diagnostic tool for assessing the quality of media as it is being produced, and critically, *before* it is fed into the cell culture reactor. To this end, using the process data that has been made available to us (protein product yield), an NPLS calibration model was built using the EEM data from the A-process Media#1 samples to attempt to identify a possible correlation of the feed media with protein product yield in g/L (data supplied by BMS). In cases where two or more ( $\geq 2$ ) EEM measurements of a sample from the triplicate runs were identified as 'normal', these normal measurements were averaged to generate a single spectrum for each sample. The resulting 27 averaged EEM measurements were split into a calibration set and a stringent validation set in a Monte Carlo sampling manner.<sup>61,62</sup> This type of analysis is difficult when the sample size is small and spectral variable size is quite large. Therefore, for robustness, five different splits of calibration and validation samples were made (labeled as follows C24/V3, C23/V4, C22/V5, C21/V6, and C20/V7, Table 3).<sup>63,64</sup> Since PLS regression must incorporate an appropriate number of LVs this training was repeated  $10^4$  times on each split to determine the most suitable model complexity using up to 15 LVs. The root mean squared error of calibration (RMSEC) and validation (RMSEV) for each calculation as function of the number of LVs were averaged and compared (Figure 4A). The appropriate model complexity was chosen according to the magnitude of these averaged RMSEVs. It is clear from Figure 4A that at LVs=5, the averaged RMSEV values from the various splits reach the global minima, which suggests that using four LVs should lead to an appropriate NPLS model. Table 3 gives the averaged RMSEC and RMSEV values obtained from the  $10^4$  calculations on each split, when four LVs were selected to yield the NPLS model. It is noticeable that the averaged RMSEV values are dependent upon validation set size, increasing as the number of measurements in the validation set increase. This phenomenon is quite common in latent variable regression<sup>65</sup> because of the small size of the sample set. The C22/V5 split was

chosen for validation as this was considered to provide the most appropriate balance of calibration and validation sets (Table 3).

**Table 3:** Determination of the Media#1 NPLS model complexity and accuracy determined by five cross-validation splits of the EEM data.

| <b>Cross-validation</b> | <b>Splits (calibration set/validation set)</b> |               |               |               |               |
|-------------------------|--|---------------|---------------|---------------|---------------|
|                         | <b>C24/V3</b>                                  | <b>C23/V4</b> | <b>C22/V5</b> | <b>C21/V6</b> | <b>C20/V7</b> |
| # Calibration samples   | 24   | 23            | 22            | 21            | 20            |
| # Validation samples    | 3  | 4             | 5             | 6             | 7             |
| # Latent variables, LV  | 4  | 4             | 4             | 4             | 4             |
| Averaged RMSEC          | 0.015  | 0.014         | 0.014         | 0.013         | 0.012         |
| Averaged RMSEV          | 0.064  | 0.065         | 0.066         | 0.066         | 0.067         |



**Figure 4:** (A) Results of cross-validation for latent variables selection to yield appropriate NPLS model for the averaged EEM data of 27 Media#1 samples; (B) NPLS model for the correlation between Media#1 EEM data and end protein product yield. Calibration model used a set of 22 randomly selected averaged Media#1 EEM spectra and validated with the 5 remaining EEM spectra.

Consequently, an NPLS calibration model was typically built using a set of randomly selected averaged EEM data from 22 Media#1 samples. Autoscaling was carried out on both the EEM spectral data and protein product yield values prior to modeling. NPLS model accuracy (Table 4 and Figure 4B) is good as indicated by the high correlation coefficient ( $R^2=0.94$ ) and low RMSEC (0.017 g/L). Note that  $R^2$  is the coefficient of correlation between the predicted and actual/measured yields, and is generally used for evaluating the model quality. Five samples were used for validation and the agreement between measured and predicted (in brackets) values in g/L for the Media#1 samples is very good: 0.79(0.84); 0.85(0.81); 0.76(0.75); 0.77(0.76); and 0.87(0.86). Replicating the NPLS analysis for the other medias stored for longer periods gave very similar results in terms of averaged RMSEV / RMSEC values and the number of LVs

(Table 4). It would appear that the minor storage induced compositional changes in these media samples (as identified by NPLS-DA) lead to reduced accuracy of their respective NPLS models. Overall prediction model accuracy is very good with respect to end protein product yield and this demonstrates the efficacy of the NPLS method. We note that when the same samples were analyzed by conventional Raman spectroscopy (see Supplemental Information), no equivalent correlations with product yield were observed. This indicates that the differences in media composition that influence product yield are very small. We also recognize that while many other factors influence the protein product yield, the high degree of correlation found with these EEM measurements does suggest that it is these small changes in the feed media which influence fluorescence emission which have a significant effect on product yield.

**Table 4:** Cross-validation results for all four media samples classes from NPLS modeling of averaged EEM data.  $10^4$  calculations were run for each specified split in a Monte Carlo sampling manner.

| <b>Cross-validation</b> | <b>Media#1</b> | <b>Media#2</b> | <b>Media#3</b> | <b>Media#4</b> |
|-------------------------|----------------|----------------|----------------|----------------|
| <b>Sample size</b>      | 27             | 26             | 24             | 27             |
| <b>Latent variables</b> | 4              | 4              | 4              | 4              |
| <b>Split</b>            | C22/V5         | C21/V5         | C19/V5         | C22/V5         |
| <b>Averaged RMSEC</b>   | 0.014          | 0.011          | 0.010          | 0.01           |
| <b>Averaged RMSEV</b>   | 0.07           | 0.07           | 0.08           | 0.09           |



## **CONCLUSIONS.**

We have demonstrated that EEM fluorescence spectroscopy coupled with multiway chemometric methods is an effective and inexpensive method for the qualitative and quantitative characterization of cell culture feed media. First, MROBPCA shows that the EEM spectra are reproducible and can be used to discriminate small changes in media blends which have obvious applicability in quality control. Second, NPLS-DA can be used to observe subtle compositional changes on prolonged storage of blended media. Finally, with NPLS we have shown that it is possible to correlate the small spectral variances in the EEM spectra with end product yield. This final result is very significant from an industrial standpoint because it can lead to significant improvements in process productivity. We are now focusing on trying to identify the specific chemical components that cause the changes in the EEM spectra, and then assess how they impact on product yield.

## **ACKNOWLEDGEMENTS**

The Irish Industrial Development Authority (IDA) and Bristol-Myers Squibb funded the Centre for Bioanalytical Sciences (CBAS), a collaborative research programme involving the National University of Ireland, Galway, Dublin City University, and Bristol-Myers Squibb. Lindy Smith (BMS) and Bryan H. Ray are thanked for their work in sample organization.

## **SUPPORTING INFORMATION AVAILABLE**

An explanation of the chemometric methods, a Raman study of these media, and other data is available. This material is available free of charge via the internet at <http://pubs.acs.org>.

## REFERENCES

- (1) Davis, J. M. *Basic Cell Culture*; 2nd ed.; Oxford University Press, 2002.
- (2) Jain, E.; Kumar, A. *Biotechnol. Adv.* **2008**, *26*, 46-72.
- (3) Wlaschin, K. F.; Hu, W. S. In *Cell Culture Engineering*; Springer: New York, 2006; Vol. 101, p 43-74.
- (4) Sitton, G.; Srienc, F. *J. Biotechnol.* **2008**, *135*, 174-180.
- (5) Nobilis, M.; Kopecký, J.; Kvetina, J.; Svoboda, Z.; Pour, M.; Kunes, J.; Holcapek, M.; Kolárová, L. *J. Pharmaceut. Biomed.* **2003**, *32*, 641-656.
- (6) Ceccato, A.; Klinkenberg, R.; Hubert, P.; Streel, B. *J. Pharmaceut. Biomed.* **2003**, *32*, 619-631.
- (7) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy, 3rd Ed*; Springer Science + Business media LLC: New York, 2006.
- (8) Patra, D.; Mishra, A. K. *Trac-Trends In Analytical Chemistry* **2002**, *21*, 787-798.
- (9) de la Pena, A. M.; Mansilla, A. E.; Diez, N. M.; Gil, D. B.; Olivieri, A. C.; Escandar, G. M. *Applied Spectroscopy* **2006**, *60*, 330-338.
- (10) Ohno, T.; Amirbahman, A.; Bro, R. *Environ. Sci. Technol.* **2008**, *42*, 186-192.
- (11) Pons, M.-N.; Bonté, S. L.; Potier, O. *J. Biotechnol.* **2004**, *113*, 211-230.
- (12) Lapworth, D. J.; Gooddy, D. C.; Butcher, A. S.; Morris, B. L. *Appl. Geochem.* **2008**, *23*, 3384-3390.
- (13) Baunsgaard, D.; Norgaard, L.; Godshall, M. A. *J. Agr. Food Chem.* **2000**, *48*, 4955-4962.
- (14) Ji Ji, R. D.; Cooper, G. A.; Booksh, K. S. *Anal. Chim. Acta* **1999**, *397*, 61-72.
- (15) Nakamura, R.; Tanaka, Y.; Ogata, A.; Naruse, M. *Anal. Chem.* **2009**, *81*, 5691-5698.
- (16) Ryder, A. G. *J. Fluoresc.* **2004**, *14*, 99-104.
- (17) Sikorska, E.; Gorecki, T.; Khmelinskii, I. V.; Sikorski, M.; De Keukeleire, D. *Food Chem.* **2006**, *96*, 632-639.
- (18) Sikorska, E.; Glisuzynska-Swiglo, A.; Insinska-Rak, M.; Khmelinskii, I.; De Keukeleire, D.; Sikorski, M. *Anal. Chim. Acta* **2008**, *613*, 207-217.
- (19) Richardskortum, R.; Rava, R. P.; Petras, R. E.; Fitzmaurice, M.; Sivak, M.; Feld, M. S. *Photochem. Photobiol.* **1991**, *53*, 777-786.
- (20) Zuluaga, A. F.; Utzinger, U.; Durkin, A.; Fuchs, H.; Gillenwater, A.; Jacob, R.; Kemp, B.; Fan, J.; Richards-Kortum, R. *Appl. Spectrosc.* **1999**, *53*, 302-311.
- (21) Baker, A. *Environ. Sci. Technol.* **2001**, *35*, 948-953.
- (22) Chen, W.; Westerhoff, P.; Leenheer, J. A.; Booksh, K. *Environ. Sci. Technol.* **2003**, *37*, 5701-5710.
- (23) Her, N.; Amy, G.; McKnight, D.; Sohn, J.; Yoon, Y. M. *Water Res.* **2003**, *37*, 4295-4303.
- (24) Wang, Z. W.; Wu, Z. C.; Tang, S. J. *Water Res.* **2009**, *43*, 1533-1540.
- (25) Hall, G. J.; Clow, K. E.; Kenny, J. E. *Environ. Sci. Technol.* **2005**, *39*, 7560-7567.
- (26) Hall, G. J.; Kenny, J. E. *Anal. Chim. Acta* **2007**, *581*, 118-124.

- (27) Lindemann, C.; Marose, S.; Nielsen, H. O.; Scheper, T. *Sensor. Actuat. B-Chem.* **1998**, *51*, 273-277.
- (28) da Silva, J.; Leitao, J. M. M.; Costa, F. S.; Ribeiro, J. L. A. *Anal. Chim. Acta* **2002**, *453*, 105-115.
- (29) Olivieri, A. C.; Arancibia, J. A.; de la Pena, A. M.; Duran-Meras, I.; Mansilla, A. E. *Anal. Chem.* **2004**, *76*, 5657-5666.
- (30) Trevisan, M. G.; Poppi, R. J. *Anal. Chim. Acta* **2003**, *493*, 69-81.
- (31) Persson, T.; Wedborg, M. *Anal. Chim. Acta* **2001**, *434*, 179-192.
- (32) Stedmon, C. A.; Bro, R. *Limnol. Oceanogr.-Meth.* **2008**, *6*, 572-579.
- (33) Rinnan, A.; Andersen, C. M. *Chemometr. Intell. Lab.* **2005**, *76*, 91-99.
- (34) Zepp, R. G.; Sheldon, W. M.; Moran, M. A. *Marine Chemistry* **2004**, *89*, 15-36.
- (35) Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A. *J. Chemometr.* **2006**, *20*, 99-105.
- (36) Eigenvector Research Inc.: 3905 West Eaglerock Drive, Wenatchee, WA.
- (37) Mathworks Inc.: Cambridge, MA, 1994-2008.
- (38) Hubert, M.; Rousseeuw, P. J.; Vanden Branden, K. *Technometrics* **2005**, *47*, 64-79.
- (39) Hubert, M.; Engelen, S. *Bioinformatics* **2004**, *20*, 1728-1736.
- (40) Joliffe, I. T. *Principal Component Analysis*; Springer-Verlag: New York, 1986.
- (41) Jackson, J. E.; Mudholkar, G. S. *Technometrics* **1979**, *21*, 341-349.
- (42) Li, G.; Chen, Z. *J. Am. Stat. Assoc.* **1985**, *80*, 759-766.
- (43) Croux, C.; Ruiz-Gazen, A. In *COMPSTAT 1996, Proceedings in Computational Statistics*; Prat, A., Ed.; Physica-Verlag: Heidelberg, 1996, p 211-217.
- (44) Hubert, M.; Rousseeuw, P. J.; Verboven, S. *Chemometr. Intell. Lab.* **2002**, *60*, 101-111.
- (45) Rousseeuw, P. J.; Driessen, K. v. *Technometrics* **1999**, *41*, 212-223.
- (46) Zuo, Y.; Cui, H.; He, X. *Ann. Stat.* **2004**, *32*, 189-218.
- (47) Liu, R. Y.; Parelius, J. M.; Singh, K. *Ann. Stat.* **1999**, *27*, 783-840.
- (48) Wold, S.; Trygg, J.; Berglund, A.; Antti, H. *Chemometr. Intell. Lab.* **2001**, *58*, 131-150.
- (49) Barker, M.; Rayens, W. *J. Chemometr.* **2003**, *17*, 166-173.
- (50) Wold, S.; Sjöström, M.; Eriksson, L. *Chemometr. Intell. Lab.* **2001**, *58*, 109-130.
- (51) Indahl, U., G.; Martens, H.; Næs, T. *J. Chemometr.* **2007**, *21*, 529-536.
- (52) Nocairi, H.; Mostafa Qannari, E.; Vigneau, E.; Bertrand, D. *Computational Statistics & Data Analysis* **2005**, *48*, 139-147.
- (53) Kemsley, E. K. *Chemometr. Intell. Lab.* **1996**, *33*, 47-61.
- (54) Sjöström, M.; Wold, S.; Söderström, B. In *Pattern Recognition in Practice II*; Gelsema, E. S., Kanal, L. N., Eds.; Elsevier: Amsterdam 1986, p 486.
- (55) Stähle, L.; Wold, S. *J. Chemometr.* **1987**, *1*, 185-196.
- (56) Fisher, R. A. *Ann. Eugen.* **1936**, *7*, 179-188.
- (57) Bro, R. *J. Chemometr.* **1996**, *10*, 47-61.
- (58) Smilde, A.; Bro, R.; Geladi, P. *Multi-way analysis with applications in the chemical sciences* John Wiley & Sons, Inc. : Chichester, 2004.

- (59) Barker, H. R.; Barker, B. M. *Multivariate Analysis of Variance (MANOVA): A Practical Guide to Its Use in Scientific Decision-Making*; University of Alabama Press Birmingham, 1984.
- (60) Huberty, C. J.; Olejnik, S. *Applied MANOVA and Discriminant Analysis, 2nd edn.*; John Wiley & Sons: New Jersey, 2006.
- (61) Picard, R. R.; Cook, R. D. *J. Am. Stat. Assoc.* **1984**, *79*, 575-583.
- (62) Xu, Q.-S.; Liang, Y.-Z. *Chemometr. Intell. Lab.* **2001**, *56*, 1-11.
- (63) Shao, J. *J. Am. Stat. Assoc.* **1993**, *88*, 486-494.
- (64) Baumann, K. *TrAC Trends in Analytical Chemistry* **2003**, *22*, 395-406.
- (65) Baumann, K.; Albert, H.; von Korff, M. *J. Chemometr.* **2002**, *16*, 339-350.