Autofluorescence Final Report

INTRODUCTION

Analytical approaches to sensing cellular reduced nicotinamide adenine dinucleotide (NADH) are significant as they can provide insights into metabolic activities. A spectral detection approach can be utilized and it was recently proven that the changes captured via a methodology are due to metabolic changes. Previous papers investigate the response of UV-excited autofluorescence to a range of well-known chemicals affecting fermentation, respiration, and oxidative-stress pathways in *Saccharomyces cerevisiae* (yeast). Previous experimentation has investigated the properties of adding different chemicals together that are known to affect both different and the same pathways, in addition to introducing an analytical approach through spectral phasor plots (breaking down the spectra into a combination of imaginary and real functions). The client wishes for us to answer a few research questions: if given many different spectra, can we derive similar components that could correlate with metabolic transitions and provide more insight into them? Can principal component analysis be used? Are there machine learning methods that could be used here?

METHODS

1. Data Processing

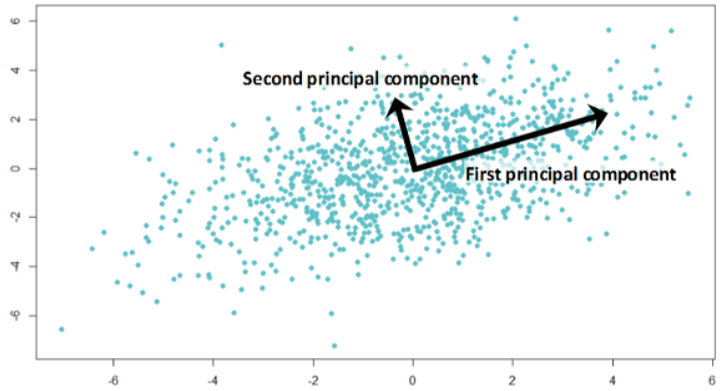
Initially, for each experiment, the estimates were recorded in multiple separate ".asc" files, with each file representing a different timestamp. For instance, the experiment titled '151211-2 propanol(OConnor)-Figure2' comprises a total of 112 distinct '.asc' files, each captured at 30-second intervals. Hence, for each experiment, we consolidate all individual '.asc' files into a single '.csv' data frame. Within the unified data frame, rows originating from the same individual '.asc' file are identified by a common ID, represented in the 'Identifier' column of the data frame. Each experiment will have a different number of substances added at several time marks. Consequently, within the consolidated data frame, we generate a 'Duration' column, based on the names of the individual '.asc' files containing timestamps, to indicate the time duration represented by each row within the experiment in seconds. Based on the 'Duration' column, we can ascertain the assignment of rows to specific substances within the experiment, with each substance being identified by a distinct ID. The values of intensity and wavelength do not display any anomalies; therefore, no data imputation is required. The processed data will resemble the following example:

	V1	V2	time	duration	experiment_id	Identifier	substance
	<dbl>	<dbl>	<dttm>	<dbl>	<dbl>	<dbl>	<dbl>
1	659.	57	2023-10-25 18:53:40	940	2	32	2
2	660.	117	2023-10-25 18:53:40	940	2	32	2
3	660.	63	2023-10-25 18:53:40	940	2	32	2
4	660.	64	2023-10-25 18:53:40	940	2	32	2
5	660.	91	2023-10-25 18:53:40	940	2	32	2
6	661.	159	2023-10-25 18:53:40	940	2	32	2

2. Starting with Principal Component Analysis (PCA)

PCA is a statistical method that is useful for simplifying a dataset with many interrelated variables while retaining the variation present in the dataset as much as possible. It is commonly used in exploratory data analysis and for making predictive models.

In a typical setting with multivariate data, PCA seeks to find the directions (called principal components) that maximize the variance in the dataset. These directions are orthogonal to each other and form a new coordinate system for the data. By projecting the original data onto these new axes, PCA provides a reduced-dimension representation where the first few axes capture the most substantial variance in the data.



- The dots on the plot represent individual data points in their new coordinates after being transformed by PCA.
- Most of the data points lie along the direction of the first principal component, indicating that this direction captures a significant amount of the variance in the data. The second principal component captures variance orthogonal to the first, but as seen, it captures less variance than the first component.

In this study, spectral data obtained from UV-excited autofluorescence of NADH can be complex, with 1024 wavelengths likely exhibiting significant multicollinearity, meaning that they change together in response to the metabolic state of the cell. PCA could be used to transform these correlated wavelengths into a set of linearly uncorrelated principal components. Each principal component would represent a combination of wavelengths that together explain a portion of the variability in the metabolic states. However, PCA is not directly applicable in this context because it assumes that the input variables are independent, while 1024 wavelengths in the spectral data are inherently related due to the nature of light absorption and emission.

3. Functional Principal Component Analysis (FPCA)

FPCA is a generalization of PCA when data is functional or in time series form. In this context, instead of looking at individual observations as points, we treat them as functions over a continuum—such as time and wavelength. Instead of examining individual values at particular wavelengths, FPCA looks at the entire spectrum as a continuous curve or function.
How FPCA works in this study:

- **Function Formation**: Each spectrum is considered a function, which might be defined over the range of wavelengths. This is an essential step because the data now reflect continuous spectral curves instead of discrete wavelength points.

- **Covariance Function**: Instead of a covariance matrix, a covariance function is estimated to describe how two points on the spectrum covary across all spectra in the sample.
- **Eigenfunctions and Eigenvalues**: Similar to PCA, FPCA involves computing eigenfunctions (analogous to eigenvectors in PCA) and their associated eigenvalues. These eigenfunctions represent the principal components in the functional space, and their eigenvalues indicate the amount of variance captured by each component.
- **Scores**: Each spectrum can be described by a set of scores corresponding to these eigenfunctions. These scores are analogous to the coefficients for principal components in PCA and indicate the weight or contribution of each eigenfunction to that particular spectrum.
- **Feature Selection**: The most significant eigenfunctions (those with the largest eigenvalues) can be selected to represent the data, reducing its dimensionality while preserving the core characteristics of the spectra.

- V1 (wavelengths, consistent throughout all data). V2 (intensity, changes). The three different spectrums are at different time ticks.

Hypotheses in the study of FPCA: In the study, hypotheses could be related to the capability of FPCA to capture and quantify the variation in spectral data due to metabolic changes:

- **Null Hypothesis (H0)**: The functional principal components derived from spectral data do not significantly represent the metabolic changes of the cells.
- **Alternative Hypothesis (H1)**: The functional principal components derived from spectral data significantly represent the metabolic changes of the cells, providing insight into the underlying biological processes.

For the research question—"Given many spectra plots, can we derive similar components that could correlate with metabolic transitions?"—FPCA is an appropriate statistical method compared to PCA. It enables us to identify whether there are common patterns (eigenfunctions) across different spectra that correspond to metabolic changes. If such patterns

are identified and found to be significant (e.g., through hypothesis testing), it would suggest that the FPCA-derived components are meaningful and could provide insights into the metabolic state of the cells.

4. Tensor Decomposition

Tensors are generalized n-dimensional vectors. A scalar is a zero-rank tensor, a vector is a one-rank tensor, and a matrix is a two-rank tensor. There exist many decomposition methods, highly reliant on linear algebra, that break down the multi-dimensional array into a sum of simpler components. Studies have shown that tensor decomposition can be used on EEG signals where each category of signal has a frequency at a given time. This can be combined into a three-dimensional tensor, which is something that mimics our current data structure.

There are various methods for tensor decomposition, with common approaches including Canonical Polyadic (CP) and Tucker decomposition. In CP, a tensor is expressed as the sum of outer products of vectors along each mode of the tensor while Tucker decomposition expresses a tensor as a core tensor multiplied by matrices along each mode (the core tensor could be something meaningful among all the spectra that we could derive and analyze). Thus, CP decomposition is simpler but more restrictive while Tucker decomposition is more flexible and can handle higher complexities.

RESULTS

I. Sequential Additions of Chemicals to Cellular samples

1. Response to ethanol (340 mM) then cyanide (8 mM).

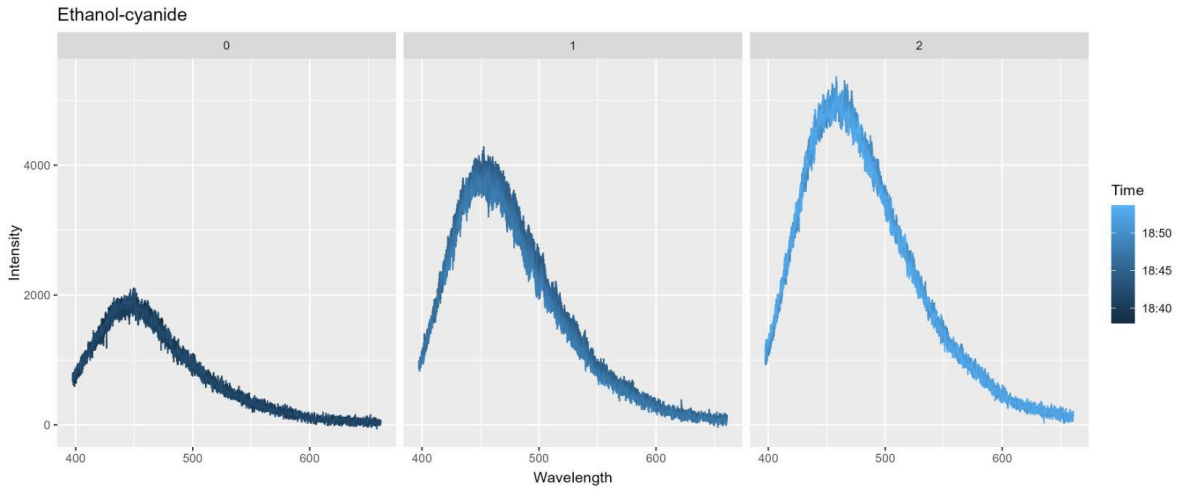
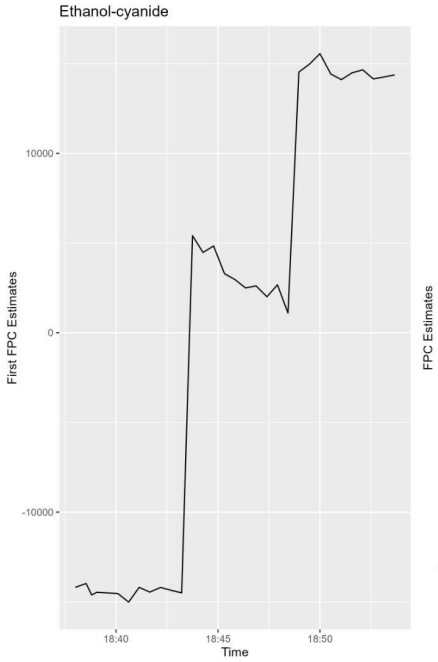


Figure 1: Autofluorescence response to the sequential additions of ethanol (340 mM) then cyanide (8 mM) over the experiment time.

According to the line graph above, there is a remarkable change in the magnitude spectra every time a new chemical is added, especially for wavelengths that are less than 500. Despite such significant growth in intensity, throughout the experiment, the autofluorescence retains its general bell shape with a peak around wavelength 450 and a heavy tail around wavelength 600.



With how distinguished the spectra change over time, only 1 functional principal component (fPC) is enough to capture 99.23% of the recorded variance. This result is consistent with the paper's finding that a two-component model cannot account for the observed change in this spectrum shape. As shown in the plot on the left, the fPC scores follow a step pattern over time, representing the high increase in intensity after each new chemical addition. However, compared to phasor analysis, FPCA falls short in displaying the distinct modes of action for ethanol and cyanide as shown in the spectral phasor shift.

Figure 2 (on the left): Estimates of fPC over the experiment time for the autofluorescence response to the sequential additions of ethanol (340 mM) then cyanide (8 mM).

2. **Response to D-glucose (added twice, 1 mM, 2 mM) then 2-deoxy-D-glucose (4 mM).**

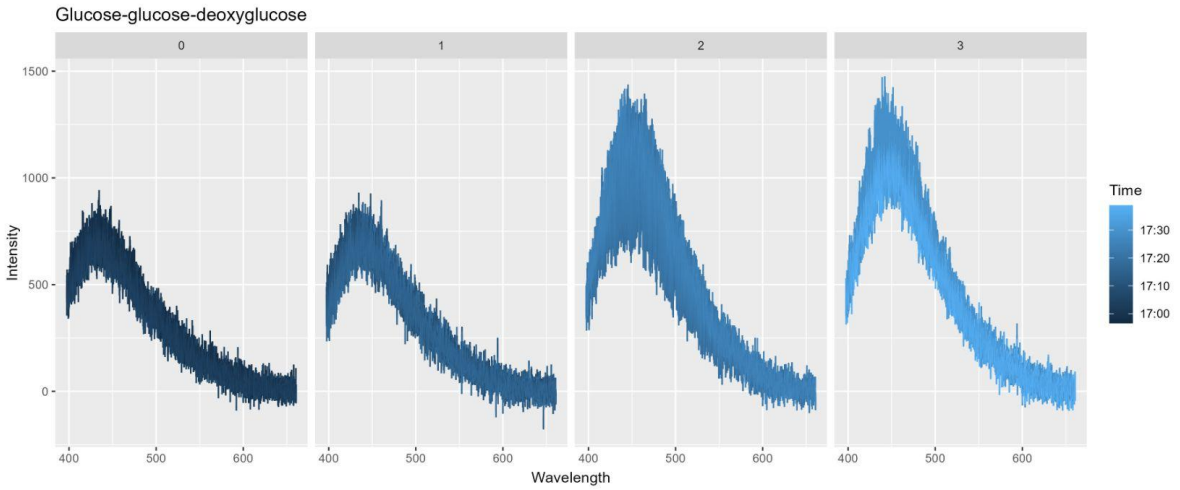
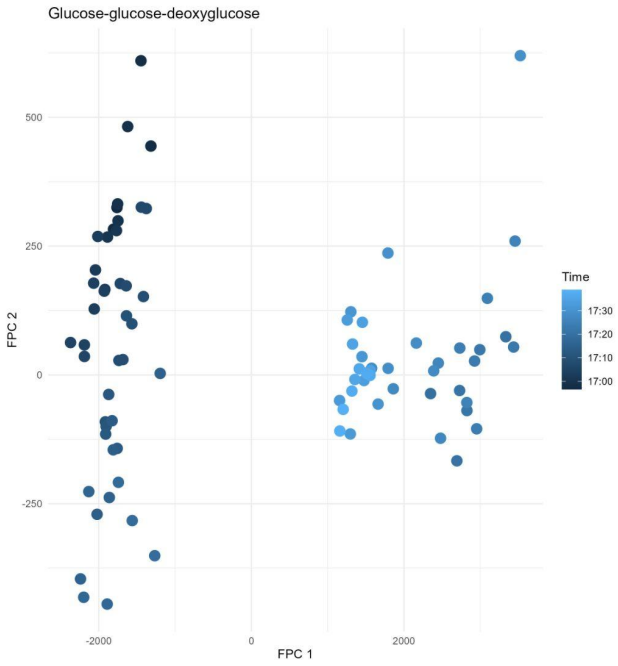


Figure 3: Autofluorescence response to the sequential additions of D-glucose (added twice, 1 mM, 2 mM) then 2-deoxy-D-glucose (4 mM) *over the experiment time.*

According to Figure 3, we can split the spectra shape change into 2 major groups: one is before and the other is after the addition of D-glucose 2mM due to their similar shape. Within each group, while the intensity was close to identical before the addition of D-glucose 2mM, the spectra fluctuated less after the addition of 2-deoxy-D-glucose.



As observed in Figure 4 here, there is a clear vertical clustering of the fPC scores over time, which aligns with the timestamp that D-glucose 2mM was added in the experiment. The first fPC explains 96.65% of the variance in this case while the second fPC only accounts for 0.01% in addition. With the first fPC capturing all the magnitude differences of the spectra, it is expected that there is no distinct clustering horizontally for the spectra after the addition of D-glucose 2mM. The slight horizontal cluster of fPC scores on the left of Figure 4 can be attributed to the occasional spikes we observed in the spectra before and after the addition of D-glucose 1mM.

Figure 4: Estimates of fPC over the experiment time for the autofluorescence response to the sequential additions of D-glucose (added twice, 1 mM, 2 mM) then 2-deoxy-D-glucose (4 mM) *over the experiment time.*

3. Response to L-glucose (4 mM), D-glucose (2 mM), then deoxy-D-glucose (4 mM)

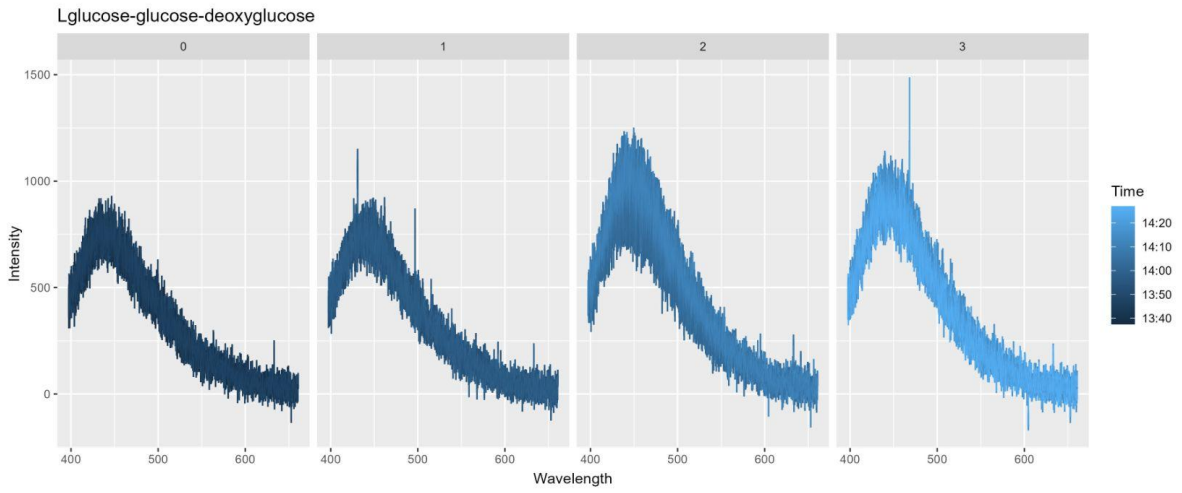
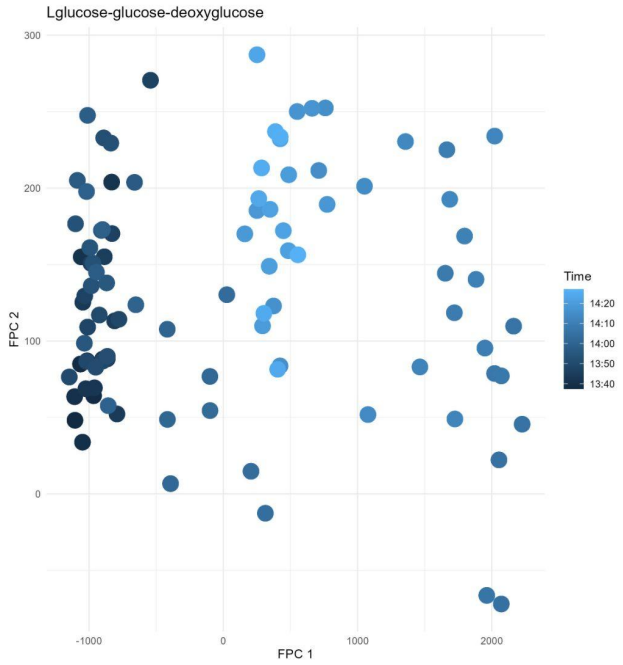


Figure 5: Autofluorescence response to the sequential additions of L-glucose (4 mM), D-glucose (2 mM), and then deoxy-D-glucose (4 mM) *over the experiment time.*

The trend repeats for this case with the spectrum can again be divided into before and after the addition of D-glucose 2mM, accompanying prominent overshooting in intensity. However, the jump in intensity is not as big compared to Figure 2 and no significant

differences in intensity magnitude can be observed within each group.



Thus, compared to the previous case, a less clear vertical clustering of the fPC scores is displayed over time here although the first FPC still explains the majority of the variance at 91.94%. However, a separation in fPC estimates is displayed here at around the mid-point of the experiment, representing the addition of D-glucose 2mM.
There is also no horizontal clustering in Figure 6, which can be blamed on how random are the outliers in this case.

Figure 6: Estimates of fPC over the experiment time for the autofluorescence response to the sequential additions of L-glucose (4 mM), D-glucose (2 mM), and then deoxy-D-glucose (4 mM) *over the experiment time.*

II. Autofluorescence response to the addition of cyanide after incubation at various d-glucose concentrations

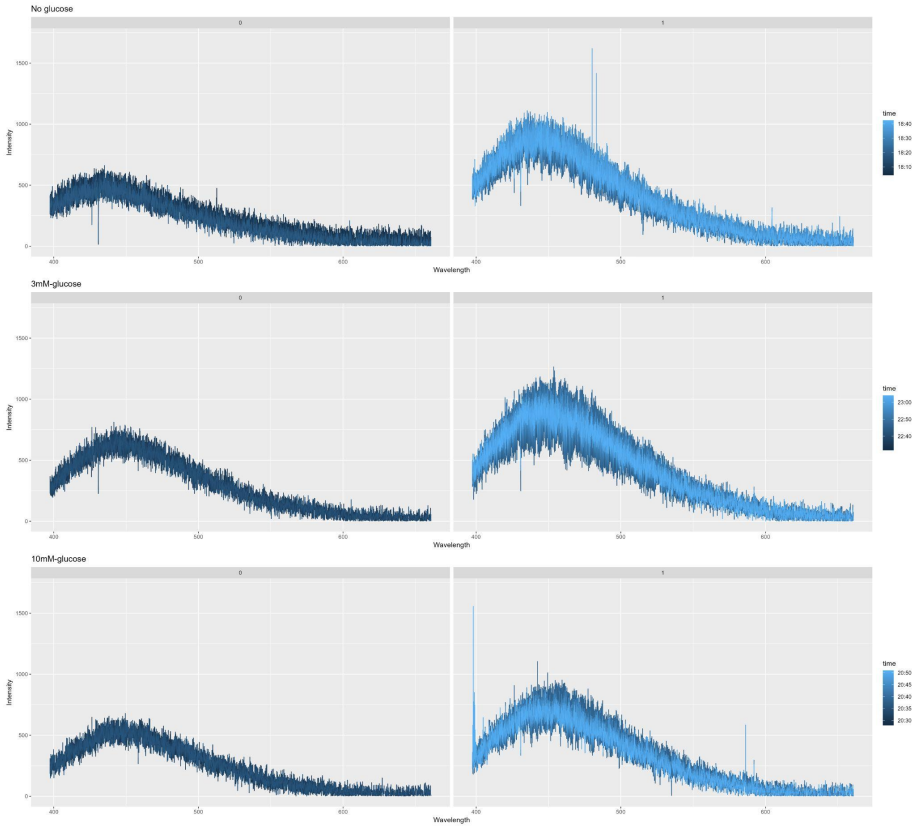


Figure 7 (previous page): Autofluorescence response to the addition of cyanide after incubation at various d-glucose concentrations. From the top down, a. Response to no glucose incubation, b. Response to 3mM glucose incubation, c.Response to 10mM glucose incubation.

The same trend is shown for all three cases in Figure 7. Whenever a new chemical is added, the intensity magnitude increases by two times at wavelengths less than 500, along with random strikes for each experiment.

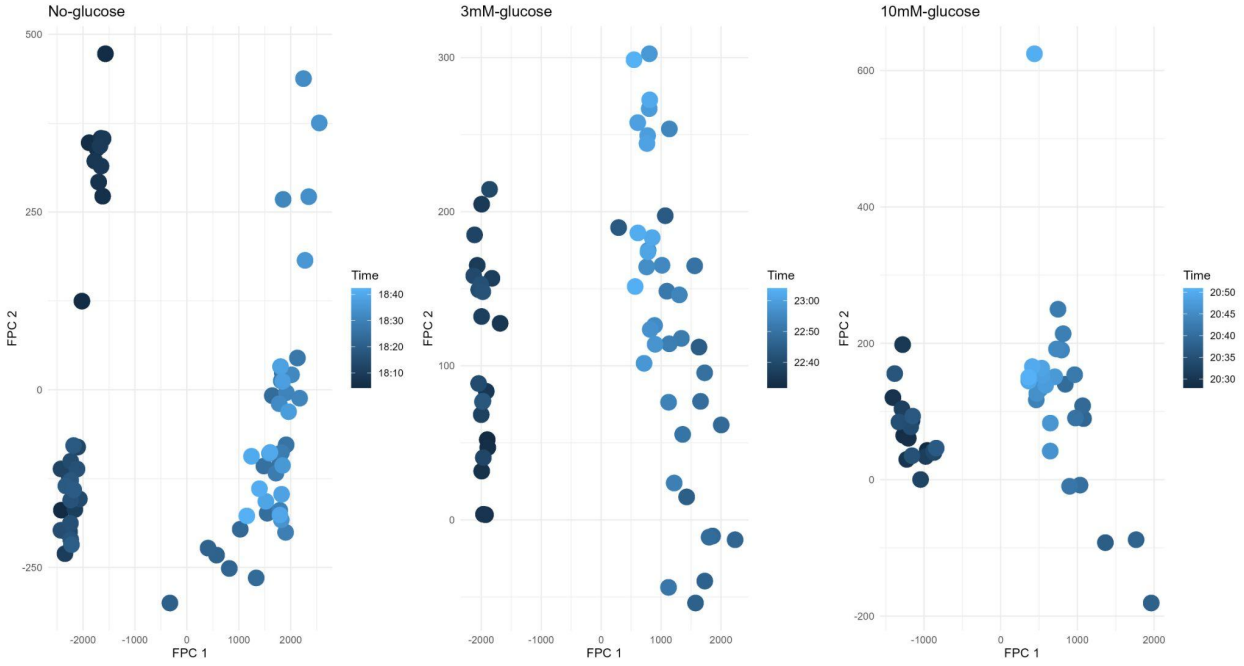


Figure 8: Estimates of fPC over the experiment time for the autofluorescence response to the addition of cyanide after incubation at various d-glucose concentrations. From left to right, a. Response to no glucose incubation, b. Response to 3mM glucose incubation, c.Response to 10mM glucose incubation.

As expected from how FPCA behaves previously with the patterned increase in intensity, a clear clustering is shown over time, corresponding to when the chemical is added for each experiment. The first FPC again explains the majority of the variance at 96.71%, 95.73%, and 90.78% for Figures 8a, 8b, and 8c respectively. Similarly, the second fPC again contributes to the explanation of outliers and noises.

III. Autofluorescence response to the addition of various alcohols (Figure 3)

1. Response to 144 mM ethanol added at 17.5 min

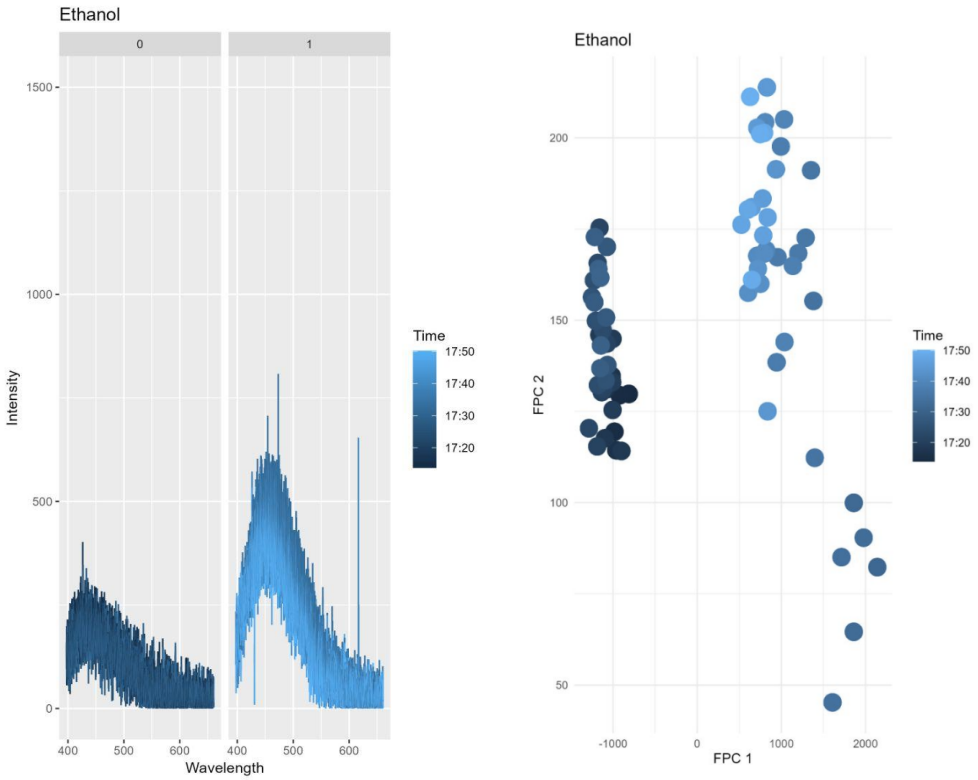


Figure 9: From left to right, a. The autofluorescence response to the addition of 144 mM ethanol at 17.5 min and b. Its estimates of fPC over the experiment time.

There is a significant increase in emission intensity after ethanol is added. Therefore, a two-component model is likely to account for the observed change. The functional principal component analysis plot reveals pronounced clustering at approximately the midpoint of the experiment, indicating the introduction of ethanol in this instance. This notable clustering primarily arises from the high variance explained by the first component, which accounts for 97% of the variance.

2. **Response to 144 mM methanol added at 7.5 min.**

As shown in Figure 10a (next page), we observe a consistent intensity pattern after the addition of methanol, despite the presence of potentially high-intensity outliers. Hence, it is improbable that a two-component model can explain the observed change. From Figure 10b, no distinct clustering is evident between the before and after addition of methanol, in contrast to previous instances. This can be attributed to the fact that the first two components explain only a minor portion of the total variance. In this instance, the first component accounts for only 33.9% of the total variance. It necessitates the inclusion of a total of 19 components to explain 95%.

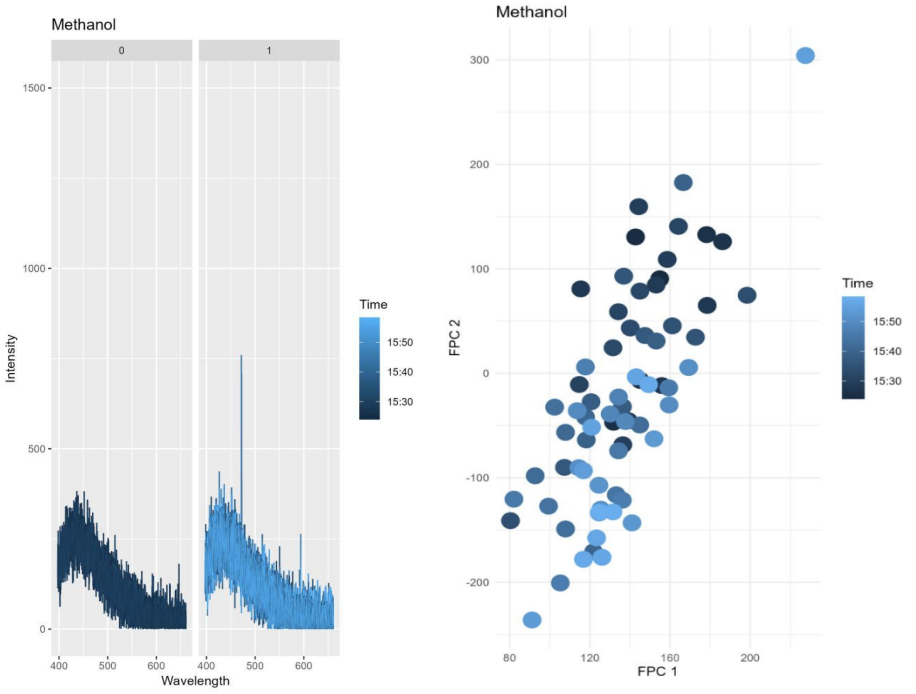


Figure 10: From left to right, a. The autofluorescence response to the addition of 144 mM methanol at 7.5 min and b. Its estimates of fPC over the experiment time.

3. Response to 144 mM 1-butanol added at 15 min

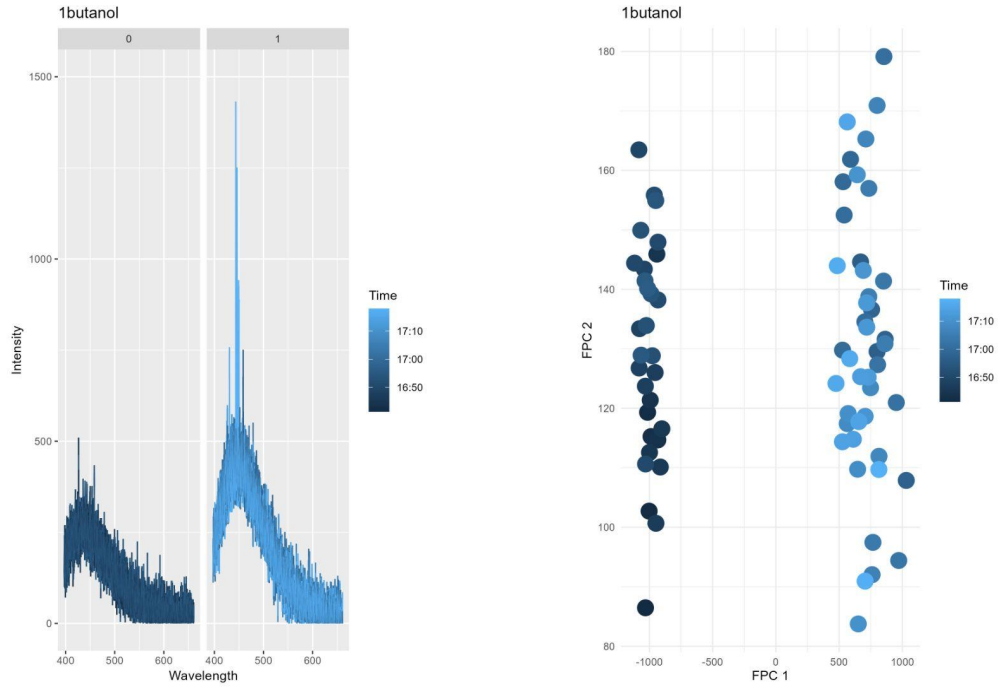


Figure 11: From left to right, a. The autofluorescence response to the addition of 144 mM 1-butanol at 15 min and b. Its estimates of fPC over the experiment time.

Based on Figure 11a, there is a significant increase in emission intensity after 1-butanol is added with extremely big outliers around wavelength 450. Therefore, similar to the ethanol case, a two-component model is likely to account for the observed change. Indeed, from Figure 11b, the functional principal component analysis plot reveals pronounced clustering at approximately the midpoint of the experiment, also indicating the introduction of 1-butanol in this experiment. This notable clustering primarily arises from the high variance explained by the first component, which accounts for 95.4% of the variance.

4. Response to 144 mM 1-propanol added at 15 min

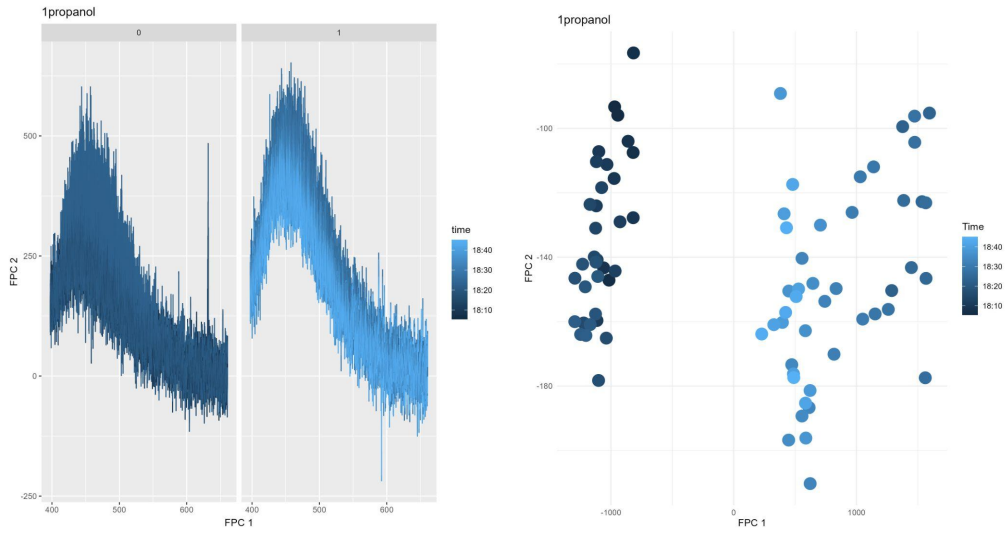


Figure 12: From left to right, a. The autofluorescence response to the addition of 144 mM 1-propanol and b. Its estimates of fPC over the experiment time.

According to Figure 12a, although there isn't a remarkable rise in emission intensity with the addition of 1-propanol, the spectra fluctuate at a narrower range, making the change in the spectra shape after the chemical addition significant. As a result, a two-component model again accounts for the observed shift with the first component's explanation for the majority of the variance, and the second fPC accounts for the noise and outliers.

5. Response to 144 mM 2-propanol added

Despite the presence of random high-intensity outliers, we see a similar intensity pattern after the addition of 2-propanol according to Figure 13a (next page), as we did with methanol. As a result, it is unlikely that a two-component model can explain the observed change. As expected, there is no discernible grouping in the spectra as shown in Figure 13b. This is because the first two components account for just a modest part of the total variation. Here, the first component accounts for only 31.96% of the overall variation in this case. It takes a total of 14 components to explain 95.59 % of the variance.

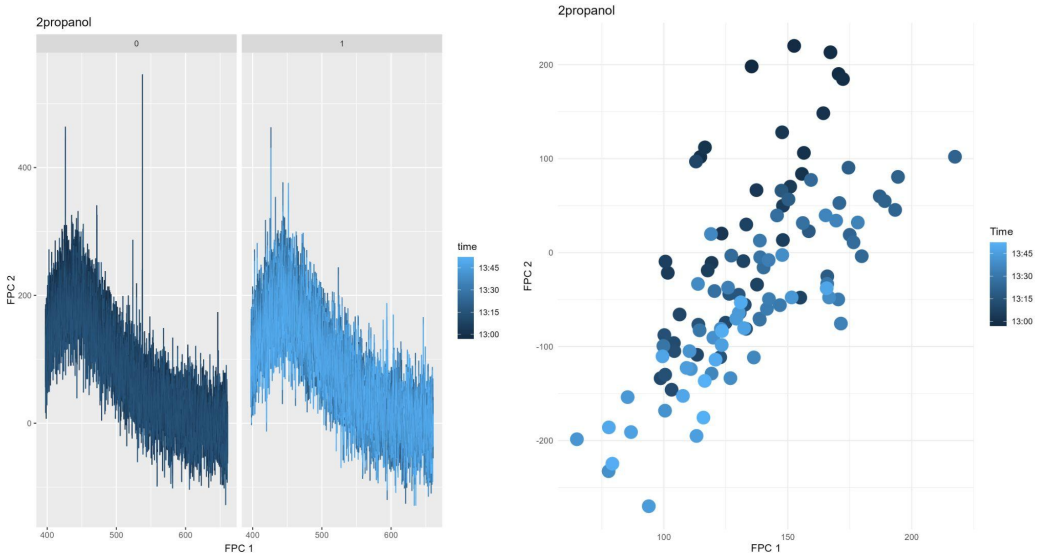


Figure 13: From left to right, a. The autofluorescence response to the addition of 144 mM 2-propanol and b. Its estimates of fPC over the experiment time.

CONCLUSION

1. Overview of Findings:

The application of Functional Principal Component Analysis (FPCA) to the spectral data in the presence of various chemicals has revealed significant insights into metabolic transitions. FPCA has enabled the dissection of complex spectral data into principal components, which has elucidated patterns and variances correlating with metabolic shifts. Notably, while a considerable amount of variability in the spectra is observed at wavelengths below 500, the functional principal component (FPC) estimates do not demonstrate marked differences when all wavelengths are considered versus when only wavelengths below 500 are considered. This suggests a degree of redundancy or non-informative variability in the higher wavelengths. The number of FPCs needed to explain the majority of variability appears to be contingent upon the magnitude of intensity changes post-chemical addition. Since most of the spectra variation can be captured through one component, the shape of the spectra themselves only changes in amplitude. Combined with our client's knowledge, this is most likely since metabolic activities were captured through the spectra, and adding chemicals only exacerbated that metabolic activity. However, when adding methanol, the first component only explained around 30% of the variation, requiring 19 components to achieve 95% variation explained. This was because methanol triggers a response in the yeast different from the other chemicals.

2. Limitations and Challenges:

One significant limitation encountered in this phase of the project is the foundational assumption that spectral data can be accurately treated as continuous functions. While this facilitates the use of FPCA, it may also gloss over some finer details intrinsic to the data, which might be important for a complete understanding of the metabolic processes.

Another challenge lies in the decision to focus on the first two FPCs. This approach assumes that these components capture the most critical response to chemical changes and the subsequent variations. However, this might result in an oversimplified view of the metabolic transitions, potentially overlooking subtle but biologically relevant changes.
The project's findings are currently limited to the effects of specific chemical additions used in the experimental design. Thus, the generalizability of these results to other metabolic stimuli or responses remains to be established.

3. Next Steps for the Project:

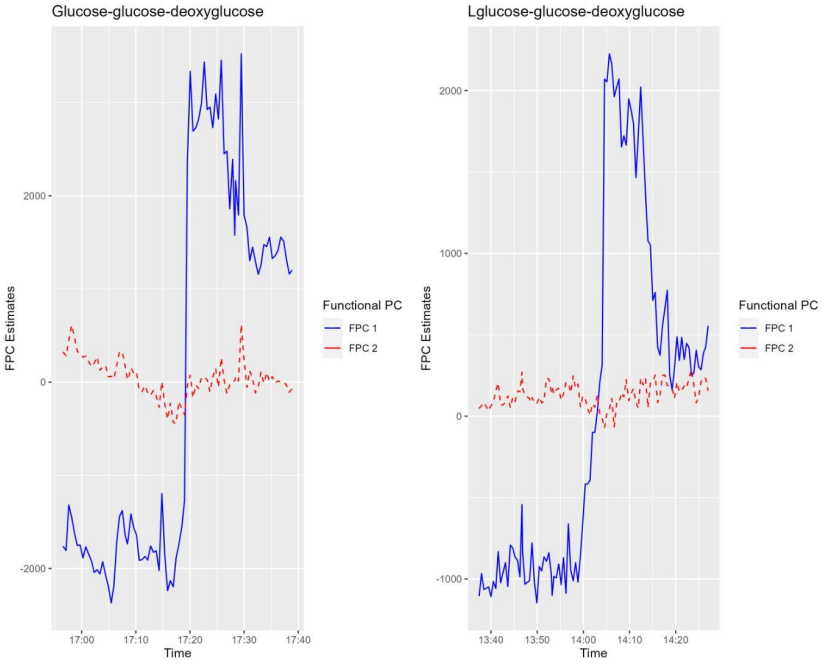
- **Further validation**: It is crucial to validate the FPCA findings by employing alternative analytical techniques, such as Canonical Polyadic and Tucker Decomposition, which might offer different perspectives on the data structure and corroborate the insights gained from FPCA.
- **Examine additional components**: Beyond the primary FPCs, analyzing subsequent components could provide a more nuanced understanding of the spectral data, ensuring that less prominent but potentially important variations are not neglected.
- **Expand chemical range**: To enhance the robustness and applicability of the findings, it would be prudent to test a broader array of chemicals and concentrations, allowing the project to assess the versatility of FPCA in capturing a wide spectrum of metabolic responses. In addition, it would be interesting to see how different chemicals with known different pathway activations would affect our components and derivations of such components.

4. Final Conclusion:

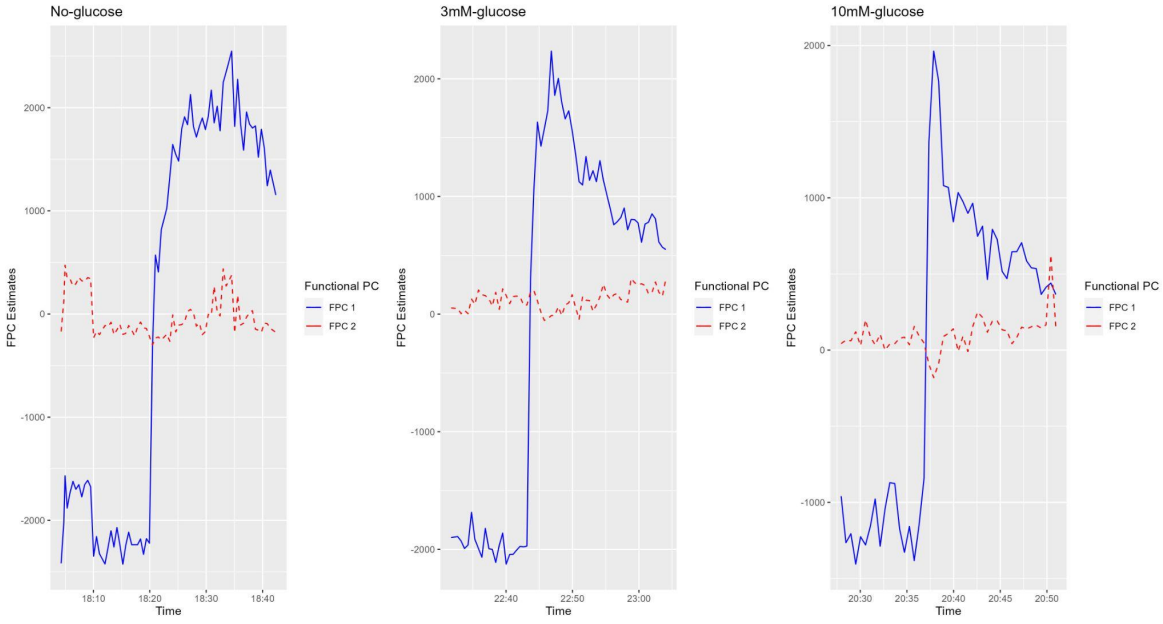
In conclusion, the current phase of the project has demonstrated the utility of FPCA in analyzing spectral data for insights into cellular metabolism. While the results are promising, the limitations acknowledged necessitate a cautious approach to interpretation and an expansive view of future research directions. The integration of additional methods and broader chemical testing will be crucial in advancing the project's objectives and enhancing the understanding of cellular metabolic processes.

APPENDIX

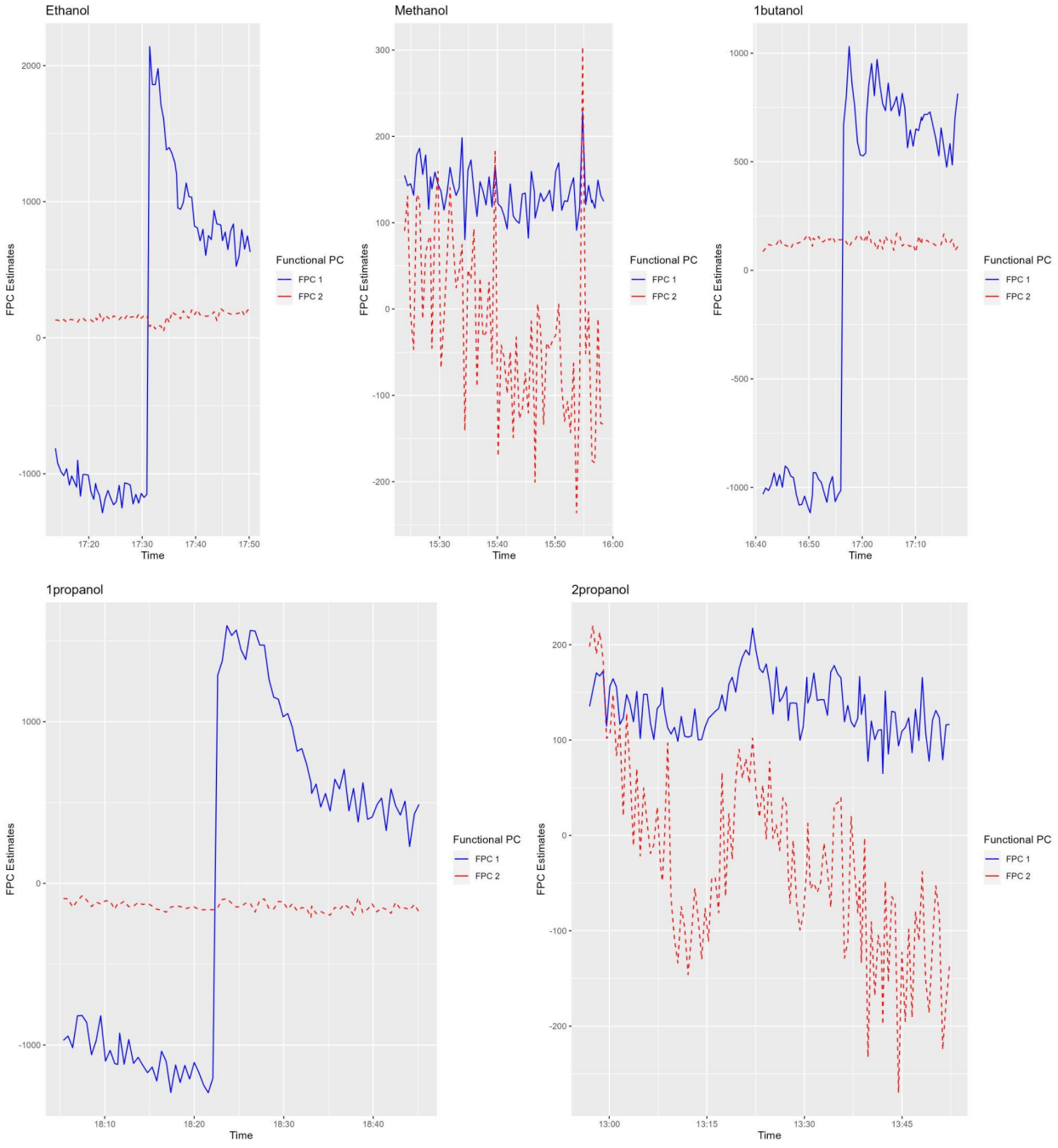
1. Sequential Additions of Chemicals to Cellular samples

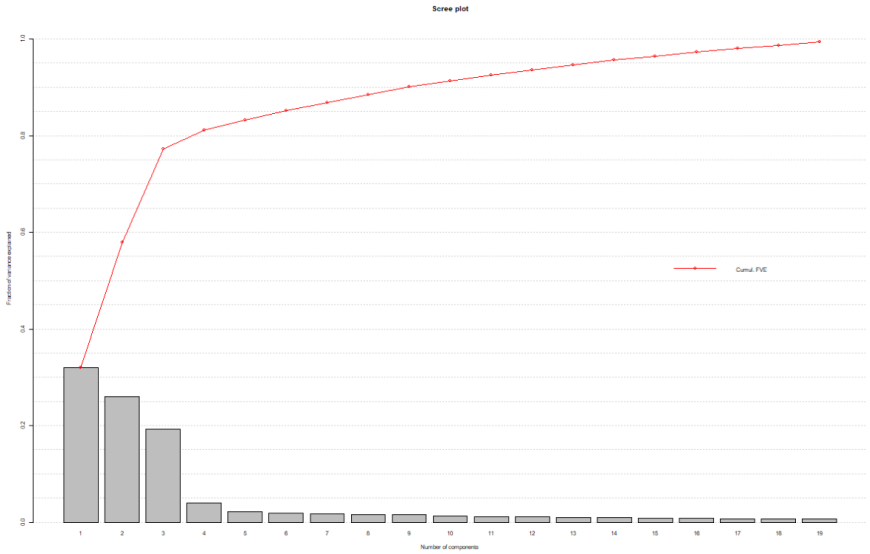


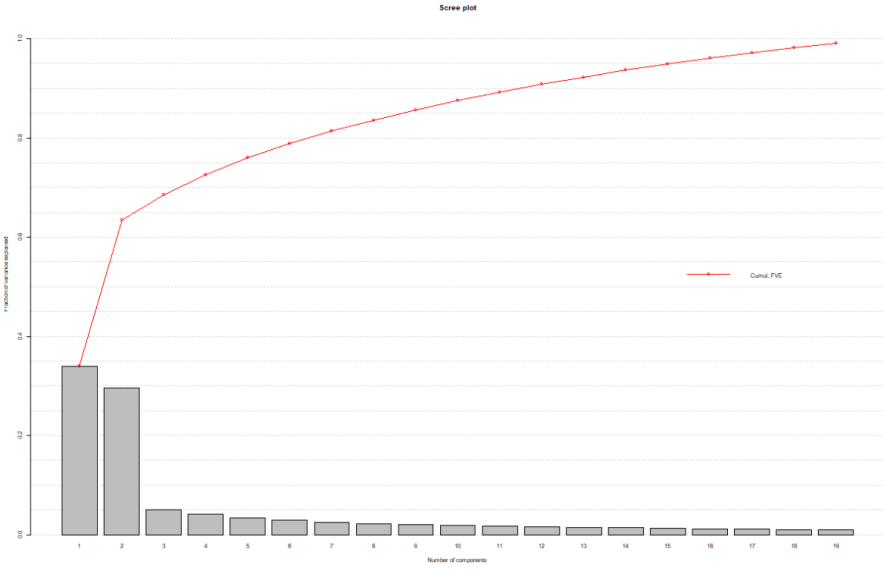
2. Autofluorescence response to the addition of cyanide after incubation at various d-glucose concentrations



3. Autofluorescence response to the addition of various alcohols (Figure 3)


Ethanol
Methanol
1butanol
1propanol
2propanol
FPC Estimates
Time
Functional PC
FPC 1
FPC 2


Scree plot
Fraction of variance explained
Number of components
Cumul. FVE


Scree plot
Fraction of variance explained
Number of components
Cumul. FVE