Functional Data Analysis for HPLC Optimization

Background and task identification

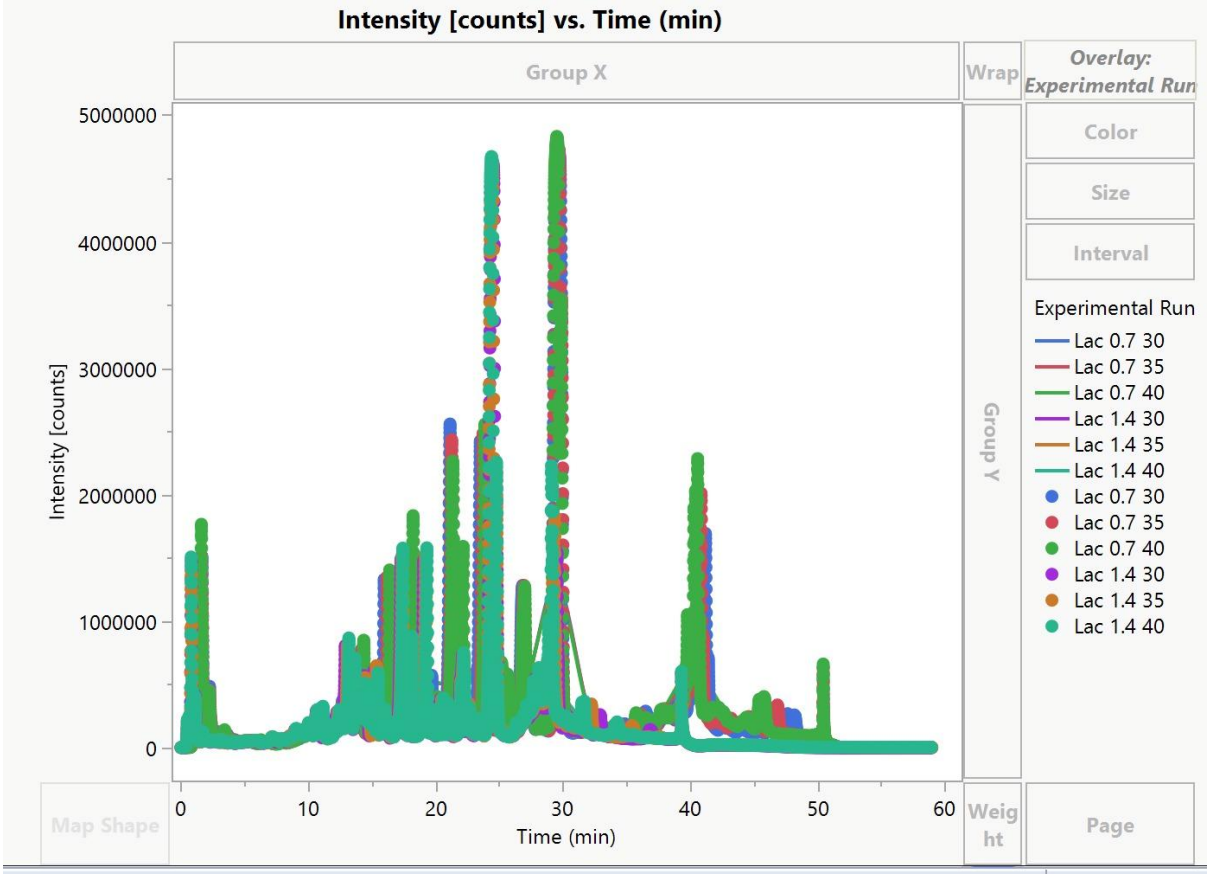
In this case study, we are confronted with a challenging analytical problem that requires the application of functional data analysis (FDA) and functional design of experiments (FDOE) to optimize an HPLC method for the accurate quantification of two closely related biological components, specifically, sophorolipid biosurfactants C18:1 and C17:1. These biosurfactants co-elute during the chromatographic process, which poses a significant challenge to accurate quantification. The objectives of this study revolve around the development of a high-performance liquid chromatography (HPLC) method that can effectively address this co-elution issue.

To achieve a suitable HPLC method, Bob needs to tackle the following key goals:

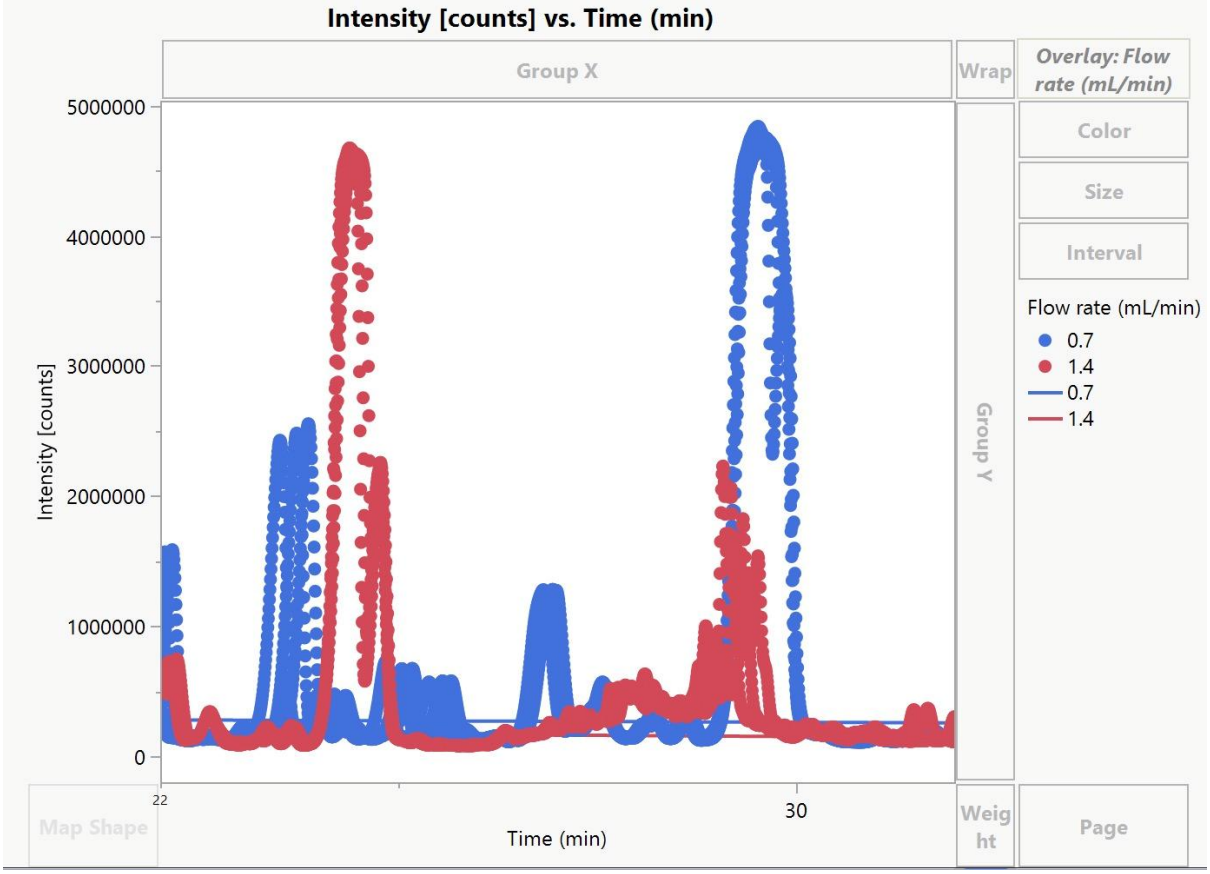
1. **Sufficient Separation of Peaks:** The primary objective is to enhance the chromatographic separation of the two closely related sophorolipid biosurfactants, C18:1 and C17:1.
2. **Peak Sharpness:** In addition to separation, Bob must ensure that the peaks are sharp, meaning they have minimal width or variance.
3. **Peak Sensitivity:** The height of the peaks is another critical aspect. Taller peaks are indicative of higher sensitivity, which means that the HPLC method can detect and quantify even low concentrations of the target analytes.
4. Use DoE to find HPLC settings that can improve the method.
5. Use FDA to understand how the curve shape changes as factors change.
6. To Identify optimized conditions for the separation of the two peaks.

Analysis & Interpretation

In this analysis, we need to separate the peaks that coelute. For that firstly, we need to interpret the values in a graph for better understanding. Analyzing the chromatographic data from various experimental runs presents a formidable challenge.

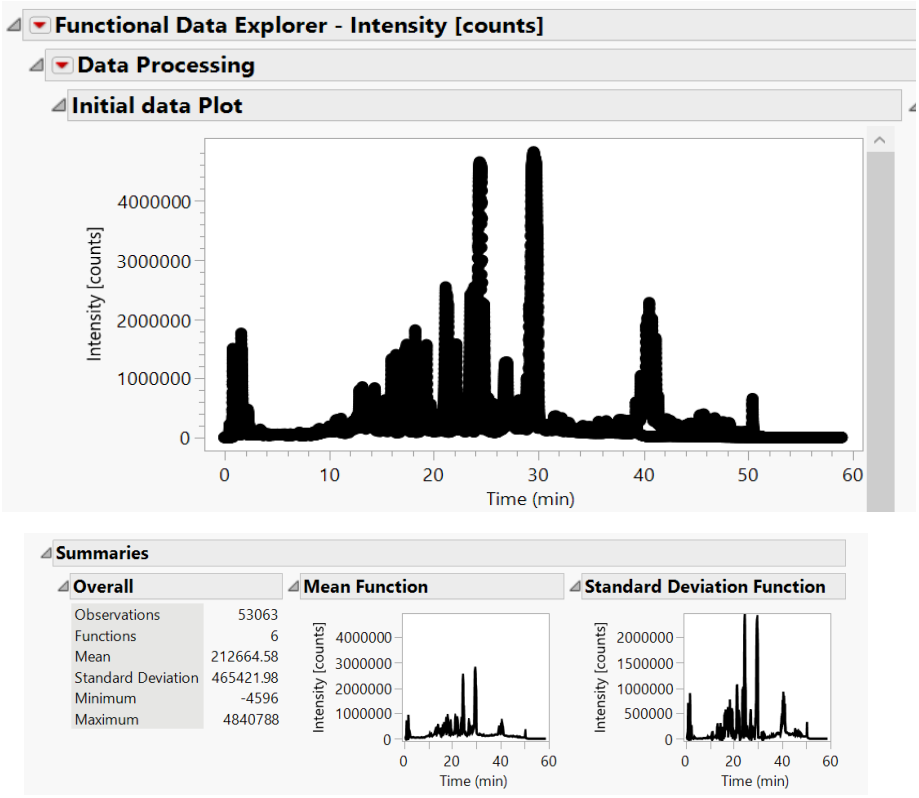


The intricacies of the data make it exceedingly intricate to discern the precise influence of temperature and flow rate variations on the chromatogram. Based on the graph, it is evident that multiple peaks are visible, and they are difficult to discern.

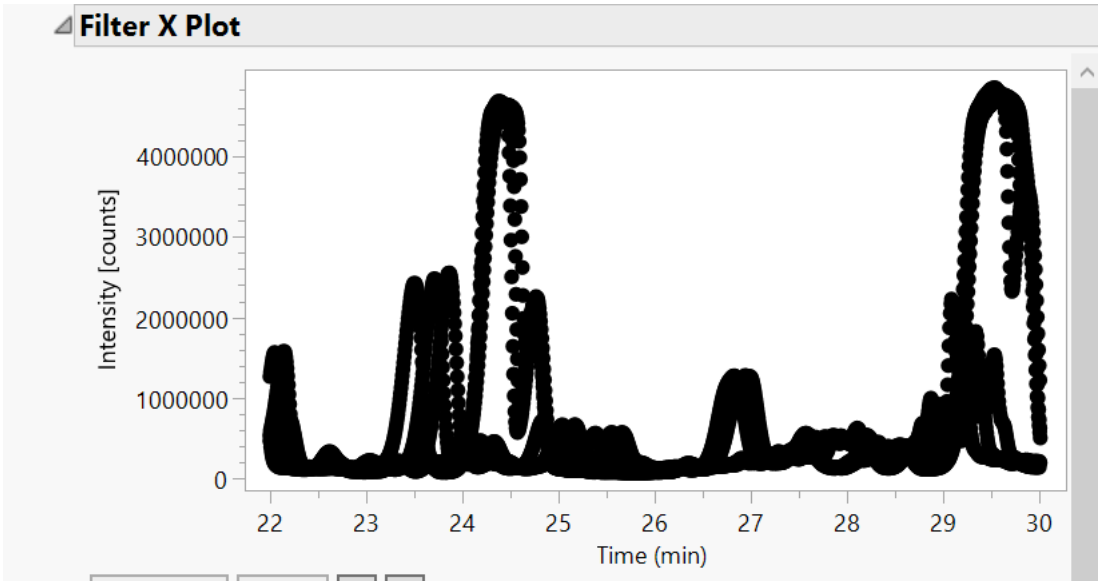


To streamline our analysis and focus on these specific peaks, it's beneficial to narrow down the X-axis range to isolate changes in this region. This allows us to zero in on the key areas of interest. Based on our previous studies with the HPLC assay, it has been consistently noted that the elution peaks corresponding to C18:1 and C17:1 tend to manifest themselves within the time frame of 22 to 31 minutes. Additionally, by grouping the data based on flow rate in each experimental run, we can identify two distinct clusters, each representing the C18:1 and C17:1 peak combination. This grouping simplifies the data and helps us distinguish the primary peaks of interest from other components. However, despite these initial observations and groupings, the overall picture remains quite complex. The chromatographic data reveals intricate relationships and interactions between the factors, such as flow rate and temperature, which affect the shape, height, and separation of the C18:1 and C17:1 peaks.

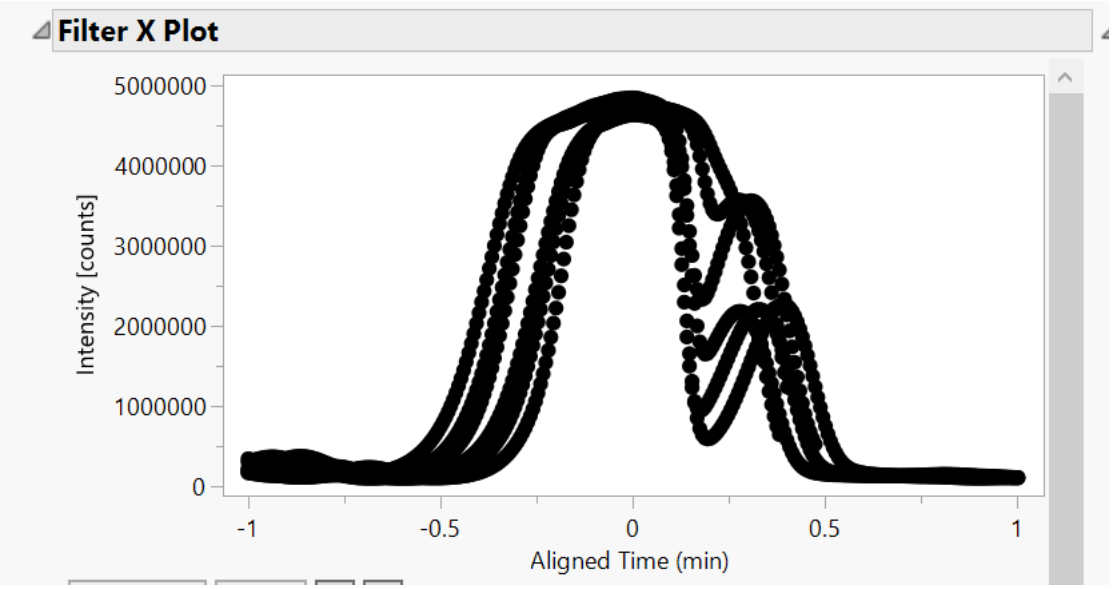
To understand how the curves evolve in response to temperature and flow rate adjustments and pinpoint the ideal settings for achieving optimal separation, we employ functional data analysis. This method allows us to delve deep into the dynamic nature of the data, providing valuable insights into the interplay between these variables.



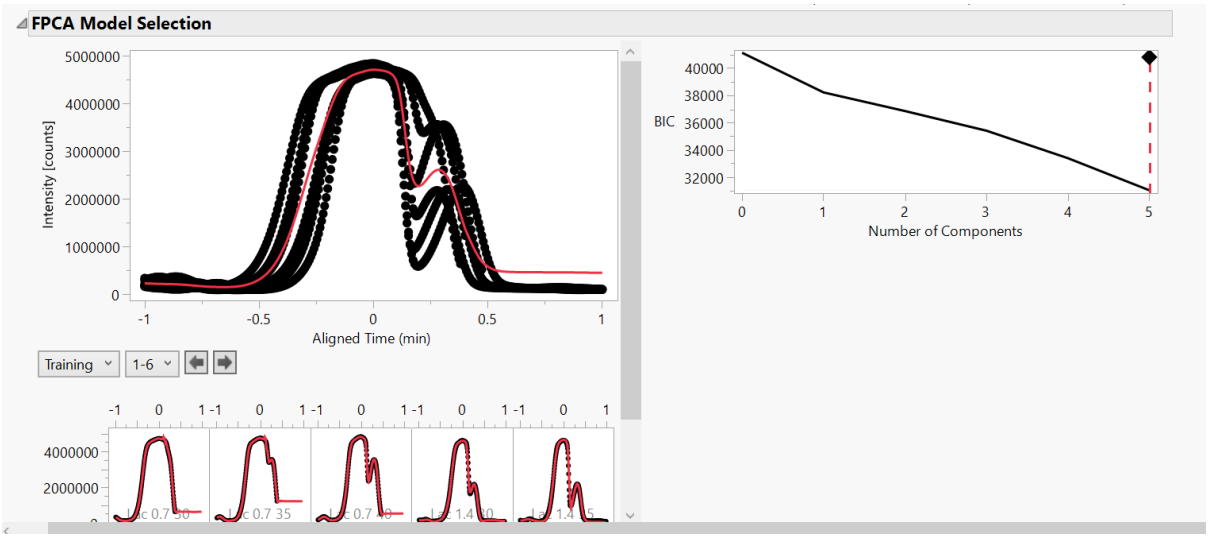
The initial data plot visually represents the dataset obtained from each HPLC experimental run, which will serve as the foundation for training the functional data model. The summaries presented here showcase both the mean function and the standard deviation function. Notably, the deviations from the mean are predominantly driven by fluctuations in temperature and flow rate. To streamline our analysis and focus on the relevant information, we narrowed down the x-axis range to 22-30 minutes. This adjustment effectively isolates the desired peak within the chromatogram, filtering out the peaks that are of lesser interest to our study.



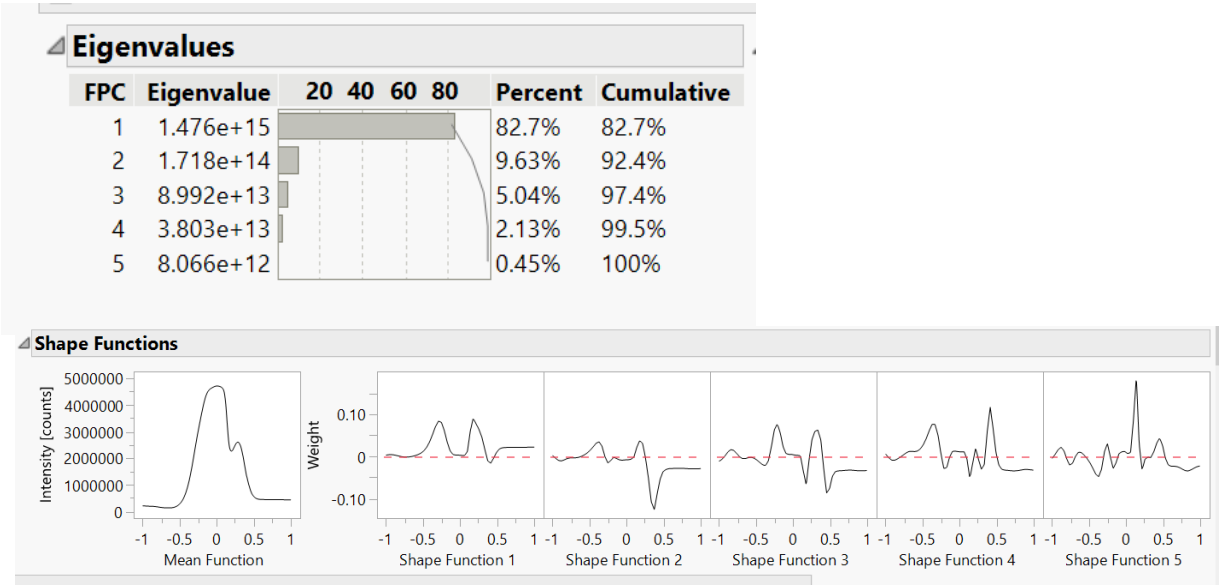
By narrowing down the x-axis range, our focus is now on the prominent peaks within each experimental run. However, it's important to note that the fluctuations in flow rate and temperature have introduced variability in the retention times of these peaks.



To isolate the specific peaks of interest, we find it necessary to perform further data refinement, effectively eliminating undesired peaks. By fine-tuning the time range to fall within -1 and 1, we successfully extract the desired peak from the chromatogram.



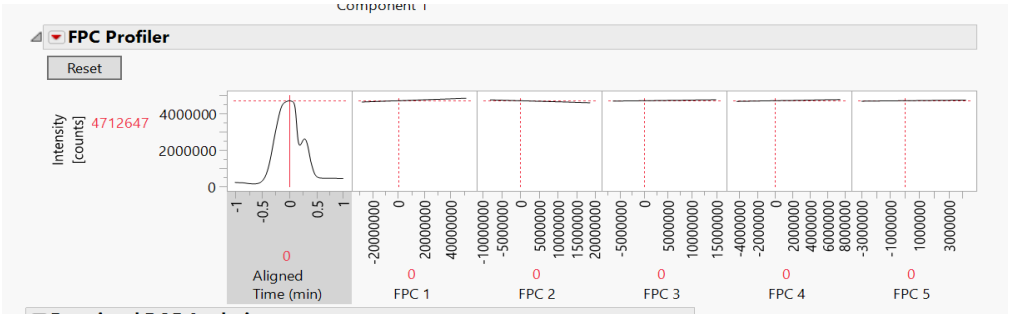
With our target peak now isolated, our primary objective is to differentiate and analyze the unique characteristics of each curve. To achieve this, we employ a smoothing model, which facilitates the transformation of our discrete, semi-continuous data into continuous functions. This approach allows us to gain a deeper understanding of the underlying patterns and variations in the chromatographic data.

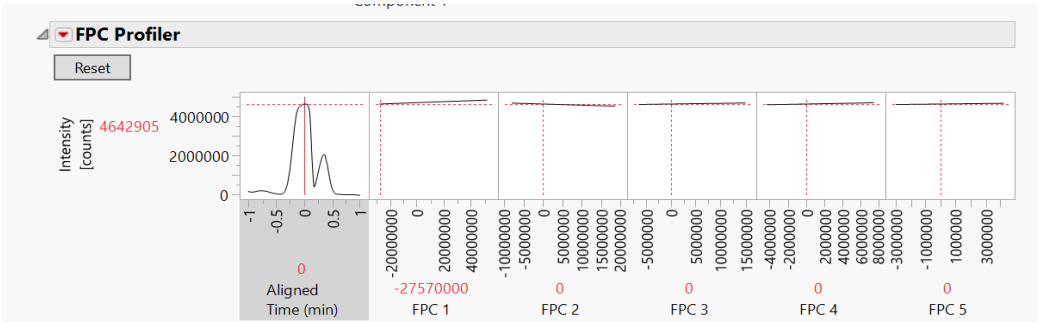


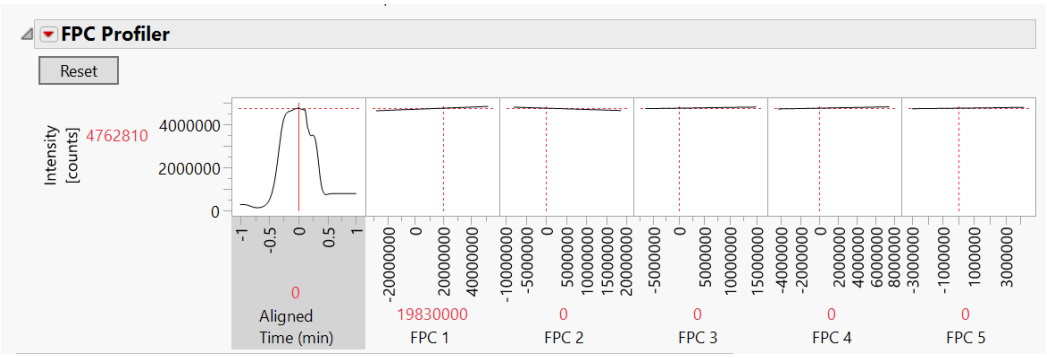
Function Summaries

Experimental Run	FPC 1	FPC 2	FPC 3	FPC 4	FPC 5	Mean	Std Dev	Median	Minimum	Maximum
Lac 0.7 30	19616205	7578278.3	-1606587	2708737.2	-331148.2	1868535.7	1817859.1	640209.47	125672.73	4823374.2
Lac 0.7 35	17172344	-5662471	-4082593	-3150004	-198268.3	1997996.2	1675547	1234801.5	128725.65	4830606.2
Lac 0.7 40	7321495.4	-6169940	6064359.4	1909627.5	735981.42	1741873.5	1748632.5	511154.43	154239.05	4838993.4
Lac 1.4 30	-7798407	6183900.2	2182692.5	-3298358	1366154	1240953.8	1616475.3	198497.39	87539.402	4626381
Lac 1.4 35	-14774851	470118.53	2315111.5	-596614.8	-2246635	1175938.1	1544245.5	216573.49	87391.081	4649544.9
Lac 1.4 40	-21536786	-2399885	-4872983	2426612.1	673916.52	1117026.6	1508893.2	212674.19	95034.159	4682237.2

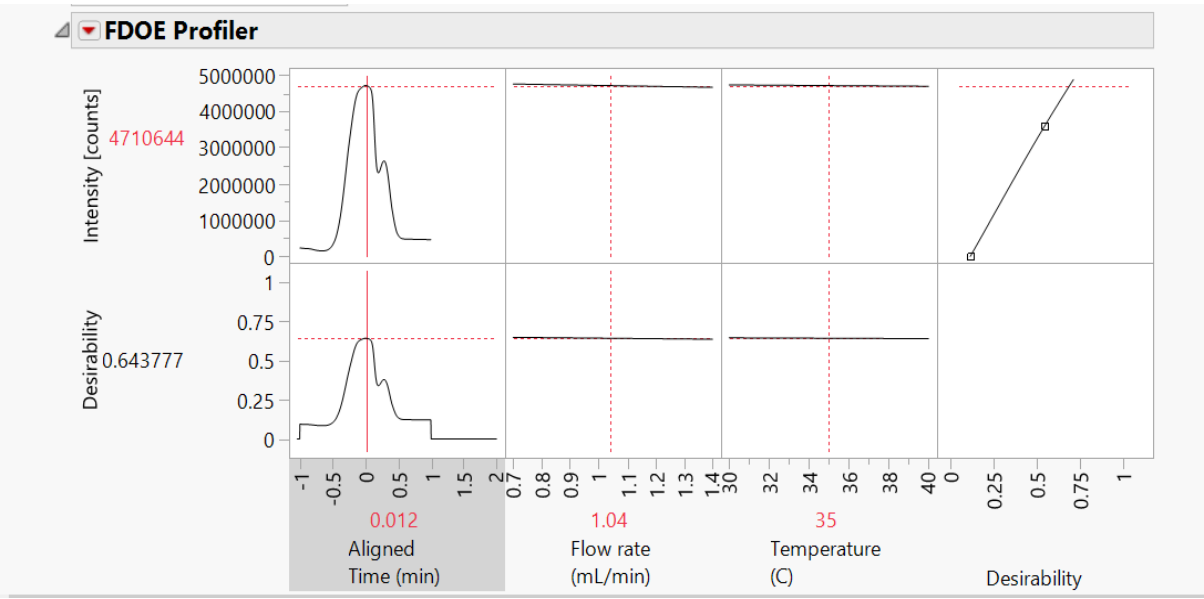
In this analysis, we employ five distinct shape functions, each serving as a representation of the prominent characteristic shapes and patterns within an experimental run. The first shape function encapsulates the most significant mode of variation, with the subsequent functions sequentially capturing the second, third, and so forth, in terms of their importance.







In this context, when the Functional Principal Component (FPC) value stands at 0, it signifies that the resulting peak closely resembles the mean function. However, as we manipulate the FPC value, either increasing or decreasing it, the resulting peak transforms, becoming wider or narrower. This leads us to the inference that the first FPC, FPC1, encapsulates the process of separating or joining the two peaks within the chromatogram. Consequently, for our desired experimental run, aiming for a low FPC value is a strategic choice, as it aligns with achieving the desired peak separation.



The primary objective of this study is to uncover the direct impact of the input parameters, such as flow rate and temperature, on the specific shape and form of the chromatogram curves. To achieve this goal, we must now explore the intricate relationships between these input variables and the resulting changes in chromatogram profiles. From the FDOE profiler, we get the maximum HPLC settings when the Flow rate is 1.04 ml/min and the Temperature is at 35 degrees Celsius. These are the optimal condition that separates two peaks.

SUMMARY

In summary, our study has employed advanced data analysis techniques to gain insights into chromatographic data. We began by refining the data to isolate the desired peaks and used smoothing models to transform semi-continuous data into continuous functions, enabling a deeper understanding of the chromatographic curves. Five shape functions were employed to represent characteristic patterns within the experimental runs, with the first function primarily capturing the most significant mode of variation. We then discussed the role of Functional Principal Components (FPCs), noting that a low FPC1 value is indicative of desirable peak separation. However, the central objective of the study is to unravel the direct influence of input parameters, such as flow rate and temperature, on the shape of chromatogram curves. This analysis seeks to establish a clear link between these inputs and the resulting chromatographic variations, contributing to a more comprehensive understanding of the system. From the analysis, the optimum settings for separating the two peaks are 35°C and 1.04mL/min. Moreover, peak separation is affected by both flow rate and temperature and sensitivity is not affected by these factors.

LIMITATIONS

The accuracy of the analysis heavily relies on the quality and consistency of the input data. Inaccurate or noisy data can introduce errors and biases in the results. Chromatographic data can be inherently complex, making it challenging to account for all possible variables and interactions that may affect the curves.

References

1. Lecture 5- PATs used in Biologics Manufacturing- 2
2. Functional Data Analysis for HPLC Optimization in JMP references Produced by Benjamin Ingham.