# *Improvement of Predictive Model to Reduce Impurity Presence Based on Raw Material Combination Pre-Selection*

*Author: William Loubriel Nieves*
*Advisor: Rafael A. Nieves-Castro, PharmD*
*Manufacturing Engineering Department*

## Abstract

Biotechnology manufacturing is composed of two main sections: Active Pharmaceutical Ingredient (API) and Drug Product (DP). The DS manufacturing most common steps are cell culture or fermentation (DSI), recovery and purification. The purification process purpose is to reduce impurities to acceptable levels. Depending on the purification raw materials, there will be different interactions with the DSI to be purified that could affect clearance capability. Multivariate analysis includes the effect of the considered variables and the response of interest. Therefore, it can be used to take into consideration combinations DSI solutions and purification raw materials available on inventory to predict the results of the impurity of interest. In this project, the improvement of a multivariate predictive model was performed by gathering recent manufacturing scale data and defining a more representative data set to improve the accuracy of the predictions. Upon implementation of the updated predictive model, several consecutive lots resulted in impurity results below the acceptable criteria.

***Key Terms - Bioprocessing, Biotechnology, Raw Materials, Prediction, Predictive Model, Purification***

## Problem Statement

While multivariate analysis can be beneficial for data processing and recommendation generation, developing an appropriate analysis can be difficult. Obtaining data to be used as foundation for the multivariate analysis can be a challenge when dealing with a large-scale manufacture. The use of laboratory scale data can generate some variability between the predicted values and the manufacturing results since the small-scale data, in some cases, is not completely representative of the large-scale process behavior. Also, large scale process improvements can result in changes that can drift the process behavior from the behavior contemplated in the analysis data set, thus creating differences between predictions and actual outcomes. Therefore, models should be updated to maintain it's currency with the manufacturing process

## Methodology

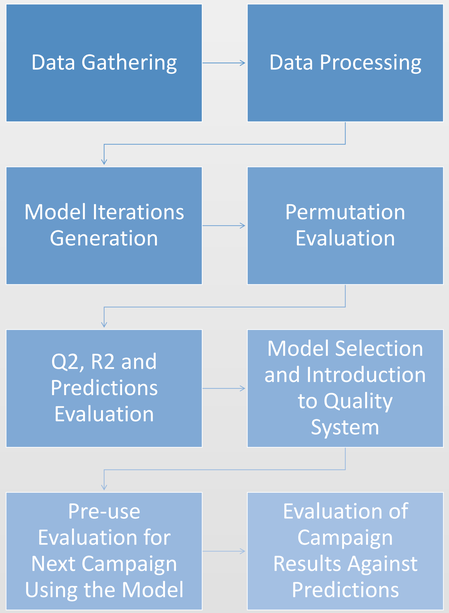
The first step for this project was process data gathering. The data was obtained from small scale runs, historical manufacturing runs, DSI intermediate characteristics and purification raw materials' certificates of analysis. The data was arranged and acclimated to the model structure so it can be used as an input. Once the data was processed, it was used for the determination of the best dataset combination.

To determine the best data set combination for model update and improvement, several predictive models iterations were generated using different datasets, this using a data analysis application. Then, each predictive model generated with different data sets was evaluated independently. The first criterion to be taken into consideration was the permutation test result. This test determined if the individual data set provide an unique solution when compared to 100 random combinations of the same data set. Models that do not pass this test were not considered as an option.

Once the models that do not comply with the permutation test were removed from the analysis, the remaining models were evaluated considering other parameters like data linearity (R2) and predictability (Q2). Most importantly, the remaining models were tested by comparing their predictions of already manufactured lots against their actual results. The manufactured lots results that were used for the test had results on the low, middle and high side of the historical results. This will helped to define accuracy of the remaining models across the spectrum of the results. The model that has the most realistic and representative data set, complies with permutation, has the best combination of R2 and Q2 and predictions was selected as the updated model.
The selected model was used to replace and update the current model. The updated model was introduced to the company official quality systems. Once the model was formally updated, it was used as part of the readiness activities. To confirm its performance is acceptable, pre-use evaluation predictions were compared against the actual lot's results. Also, it was confirmed that the actual results comply with the impurities acceptance criteria.

### METHODOLOGY

**Figure 1: Methodology Strategy**




**Project Timeline**

**Table 1: Proposed Project Timeline**

| Activity | Timeline |
|---|---|
| Data Gathering and Data Processing | August |
| Model Iterations Generation, Models Evaluation and Selection | September-November |
| Model introduction to Quality Systems | November |
| Pre-use evaluation | December |
| Products Campaign | January -February |
| Evaluation of Campaign Results | March-April |

## Results and Discussion

The historical data was gathered for manufactured lots from year 20XX to present. After data was gathered, it was properly processed to make it fit for use inside the model. The original data set was composed of xx lots/runs between small scale and full-scale manufacture data. For the model update, xx historical lots were considered. From all xx lots a total of 10 potential predictive models were generated using data analysis program and applying multivariate analysis. From the total of 10 potential predictive models, the number of options was reduced to 7 based on permutation test results. For reference, Figure 2 and Figure 3 show unacceptance and accepted permutation tests examples. Although all factors defined in the methodology were considered, the selection of the appropriate model was based on two major factors; model's data set was a better representation of the current manufacturing process, and the model showed the best predictions on the side that could trigger nonacceptable results.

**Figure 2: Unacceptable Permutation Example**

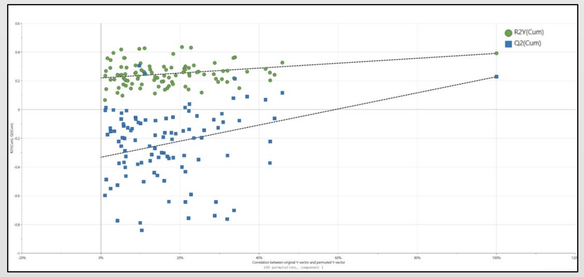
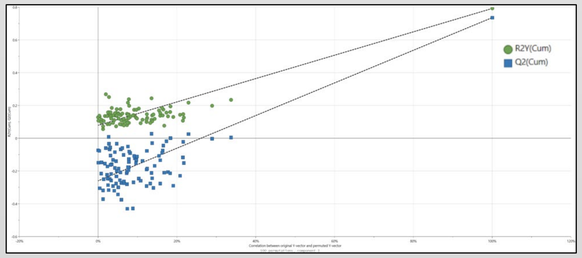



**Figure 3: Acceptable Permutation Example**

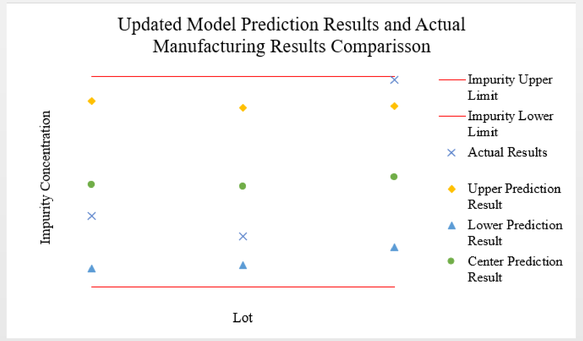



Once the model was selected, it was used to define the raw material combinations for several manufacturing lots. For this purpose, all the lots of raw materials and the DSI product lots available were crossed against each other using the updated model. More than 70 combinations options were evaluated with the new model. Based on the predictions results, the best combinations options were narrowed down to the expect campaign amount of lots plus one spare lot.

Once the lots were manufacture, the impurity results were obtained. The lots complied with the acceptance criteria as predicted by the model. When evaluated in detail, only one of the predictions was outside of the prediction range but, although close, it was within the acceptance criteria as shown in Figure 4 . This was expected since the model was not selected by its precision on the acceptable side, but for its accuracy on the non-acceptable side. Also, the updated model, although more representative, possess a more limited data set than the previous model.

Therefore, the pre use raw material evaluation using the updated predictive model was effective on reducing the probability of obtaining impurity results outside of the acceptable criteria. This results on a successful campaign with 100% success rate, product availability and cost avoidance on resources to manage deviations.

**Figure 4: Comparison of Lower, Center and Upper Prediction Results against Actual Manufacturing Results for Three Lots**




## Conclusion

The purpose of this project was to update a raw material pre-use predictive model to generate impurity predictions that could reduce the probability of obtaining impurities results outside of the acceptable criteria. Upon implementation of the updated predictive model, the results were satisfactory since consecutive lots obtained impurity results within the acceptable criteria. Consequently, no deviation was generated, and product will be available for distribution to patients.
One of the limiting factors for this project was data availability. The updated model, although more representative, possess a more limited data set than the previous model. Manufacture data can only be obtained from completed lots. Therefore, the data can be improved at the same pace as the lots are manufacture. For future work and once more data is generated, the model can include these new lot's data to keep adding robustness to its predictions and continue having satisfactory results. It is also recommended the automation of the data gathering, processing and model generation to reduce workload

## References

1. Flickinger, M. C. (2013). Downstream Industrial Biotechnology: Recovery and Purification. John Wiley & Sons, Inc.
2. Guo, L., Zhang, B., & Zhao, X. (2021). A Consumer Behavior Prediction Model Based on Multivariate Real-Time Sequence Analysis. Mathematical Problems in Engineering, 1-5. Retrieved from https://ezproxy.pupr.edu:2093/10.1 155/2021/6688750
3. JMP Statistical Discovery LLC. (April 5, 2023). Retrieved from JMP Learning Library Web site:https://www.jmp.com/en_nl/learning-library/to pics/multivariate-methods.html
4. Mengual-Macenlle, N., Marcos, P. J., Golpe, R., & González-Rivas, D. (2015). Multivariate analysis in thoracic research. Journal of thoracic disease, 1-5.
5. Minitab, LLC. (April 5, 2023). Retrieved from Minitab Support Web site: https://support.minitab.com/en-us/minitab/20/help-and-how-to/statistical-modeling/multivariate/suppo rting-topics/basics/what-is-multivariate-analysis/
6. Nayana, B., Kumar, K. R., & Chesneau, C. (2022). Wheat Yield Prediction in India Using Principal Component Analysis-Multivariate Adaptive Regression Splines (PCA-MARS)., (pp. 461-474). Retrieved from https://ezproxy.pupr.edu: 2093/10.3390/agriengineering4020030
7. Sissolak, B., Lingg, N., Sommeregger, W., Striedner, G., & Vorauer-Uhl, K. (2019). Impact of mammalian cell culture conditions on monoclonal antibody charge heterogeneity: an accessory monitoring tool for process development. Journal of Industrial Microbiology & Biotechnology, 1167-1178. Retrieved from https://ezproxy.pupr. edu:2093/10.1007/s10295-019-02202-5
8. Rathore, A., Bhambure, R., & Ghare, V. (May 18, 2010). Obtenido de SpringerLink: https://link.springer.com/article/10.1007/s00216-01 0-3781-x
9. Sokolov, M., Soos, M., Neunstoecklin, B., Morbidelli, M., Butté, A., Leardi, R., . . . Broly, H. (2015). Fingerprint Detection and Process Prediction by Multivariate Analysis of Fed-Batch Monoclonal Antibody Cell Culture Data. Biotechnology Progress, 1633-1644. Retrieved from https://ezproxy.pupr.edu:2093/10.1002 /btpr. 2174